342
1.9.70

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Edited by : P. C. MAHALANOBIS

VOL. 18, PARTS 1 & 2 MAY 1957

ON THE PERFORMANCE CHARACTERISTIC OF CERTAIN METHODS OF DETERMINING CONFIDENCE LIMITS

By B. M. BENNETT

University of Washington, Seattle, USA

SUMMARY

The article examines the performance characteristic of certain statistical procedures in current use particularly in bioassay for the determination of confidence limits in the following situations :

(i) the ratio of two means from a bivariate normal population,

(ii) the abscissa corresponding to a given ordinate in a regression problem,

(iii) the intersection point of two population regression lines.

1. INTRODUCTION

A rule of dealing with statistical data is usually termed a statistical procedure. Such a procedure will frequently lead to certain conclusions concerning an unknown distribution function. In particular if H is a certain statistical hypothesis and R a proposed method of testing H, i.e., a rule by which H will be accepted or rejected accordingly as a random function $M(\mathbf{x})$ does or does not possess certain values or properties in terms of the n observations, $\mathbf{x} = (x_1, \ldots, x_n)$ on a random variable, then R is a statistical procedure.

Let θ_1 represent an unknown parameter in the distribution $f(\mathbf{x}, \boldsymbol{\theta})$ of the observed variates; here $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)$ is the vector consisting of the s parameters of the distribution. Suppose that $\underline{\theta}_1(\mathbf{x})$, $\bar{\theta}_1(\mathbf{x})$ are a certain pair of functions and the rule states that the true value of θ_1 is included in the interval $(\underline{\theta}_1, \bar{\theta}_1)$. This rule or determination of an *interval estimate* of θ_1 is a statistical procedure. If Ω is the class of all possible distribution functions f, then the probability

$$\pi(f) = P\{\underline{\theta}_1(x) \leqslant \theta_1 \leqslant \bar{\theta}_1(x) / f\}, \qquad \ldots \quad (1.1)$$

EAU OF EDUCL. & PSYL.

considered as a function of the distribution f in Ω will be termed the performance characteristic of the method of estimating the parameter θ_1 based on the functions $\underline{\theta}_1(\mathbf{x})$, $\bar{\theta}_1(\mathbf{x})$. In general the probability $\pi(f)$ may involve $\boldsymbol{\theta}$ also. If $\underline{\theta}_1(\mathbf{x})$, $\bar{\theta}_1(\mathbf{x})$ are *confidence limits* with an assigned confidence coefficient $= 1-\epsilon$, then it is required that

$$\pi(f) = P\{\underline{\theta}_1(\mathbf{x}) \leqslant \theta_1 \leqslant \bar{\theta}_1(\mathbf{x})\} = 1-\epsilon, \qquad \ldots \quad (1.2)$$

in $\boldsymbol{\theta}$ for every f in Ω.

It is known that the problem of interval estimation of one or more of the parameters of $\boldsymbol{\theta}$ is equivalent to that of determining regions of acceptance $A(\boldsymbol{\theta})$, and this duality between the problem of confidence limits and that of regions of acceptance then enables one to determine the former under certain conditions (Neyman, 1937).

2. CONFIDENCE LIMITS FOR THE RATIO OF TWO MEANS

Let (x_i, y_i), $(i = 1, \ldots, n)$ represent a sample of n observations from a bivariate normal distribution with means

$$E(x) = \xi (\xi > 0),\ E(y) = \alpha\xi,$$

and the variances and covariance $\sigma_x^2 = \sigma_1^2$, $\sigma_y^2 = \sigma_2^2$, $\sigma_{xy} = \rho\sigma_1\sigma_2$, respectively. For convenience we shall assume σ_1, σ_2, ρ to be *known*.

The problem of determining confidence limits for α is equivalent to that of finding a 'best' region of acceptance for the hypothesis: $\alpha = \alpha_0$.

We first introduce an orthogonal transformation, $\mathbf{u} = \mathbf{A}\,\mathbf{x}$, $\mathbf{v} = \mathbf{A}\,\mathbf{y}$ on each set of variates, where $\mathbf{A}$ is the Helmert matrix of n rows and columns, $\mathbf{x}$, $\mathbf{y}$ the column vectors of the original variates, and $\mathbf{u}$, $\mathbf{v}$ those of the new variates.

Concerning the mean values of the new set (u_i, v_i), $(i = 1, \ldots, n)$, we note that

$$E(u_i) = E(v_i) = 0, \quad (i = 1, \ldots, n-1),$$
$$E(u_n) = \sqrt{n}\,\xi, \quad E(v_n) = \sqrt{n}\,\alpha\xi.$$

In virtue of the orthogonality of the transformation, the variances and covariance are unchanged.

Making the further transformations :

$$u_i' = u_i/\sigma_1,\ v_i' = v_i/\sigma_2,$$

so that $E(u_n') = \xi' = \sqrt{n}\,\xi/\sigma_1, \quad E(v_n') = \alpha'\xi' = \sqrt{n}\,\alpha\xi/\sigma_2,$

the new sets (u_i, v_i), $(i = 1, \ldots, n)$ have now unit variance, and covariance $= \rho$. Henceforth we drop the primes.

The appropriate region for testing the hypothesis : $\alpha = \alpha_0$ is given by

$$|z| = \frac{|v_n - \alpha_0 u_n|}{\sqrt{1-2\rho\alpha_0+\alpha_0^2}} > t_\epsilon \qquad \ldots \quad (2.1)$$

Bureau Edn!. [illegible] Research

if t_ε is a value determined from the tables of the normal integral such that

$$\frac{1}{\sqrt{2\pi}}\int_{-t_\varepsilon}^{t_\varepsilon} e^{-\frac{1}{2}\tau^2}\,d\tau = 1-\epsilon.$$

Confidence limits for α are then available from the region : $|z| \leqslant t_\varepsilon$.

Since the region : $|z| \leqslant t_\varepsilon$ provides 'short unbiassed' (Neyman, 1937) confidence limits for α, let us consider the specific character of these limits. Regarded as a function of α, $|z| \leqslant t_\varepsilon$ yields the following inequality as a quadratic equation in α,

$$\lambda_0\alpha^2 - 2\lambda_1\alpha + \lambda_2 \leqslant 0 \qquad \ldots \quad (2.2)$$

where $\lambda_0 = u^2 - t_\varepsilon^2$, $\lambda_1 = uv - \rho t_\varepsilon^2$, $\lambda_2 = v^2 - t_\varepsilon^2$, if we omit the subscripts on u_n and v_n; furthermore denote the discriminant of the quadratic in (2.2) by

$$\Lambda = \lambda_1^2 - \lambda_0\,\lambda_2 = t_\varepsilon^2\,[(u^2 - 2\rho uv + v^2) - t_\varepsilon^2(1-\rho^2)]. \qquad \ldots \quad (2.3)$$

The roots of the quadratic equation are, say

$$\underline{\alpha} = \frac{\lambda_1 - \sqrt{\Lambda}}{\lambda_0}, \qquad \bar{\alpha} = \frac{\lambda_1 + \sqrt{\Lambda}}{\lambda_0} \qquad \ldots \quad (2.4)$$

and these presumably are the confidence limits for α, i.e., we can state that

$$P\,\{\underline{\alpha} \leqslant \alpha \leqslant \bar{\alpha}\} = 1-\epsilon.$$

Let us now consider two present versions of the 'quadratic method' of obtaining cnfidence limits for the ratio α. Consider then Fieller (1940, 1944).

Method 1. (Fieller's Theorem): *Confidence limits* $\underline{\alpha}$, $\bar{\alpha}$ *for the ratio* α *are given by* (2.4) *provided the coefficient* $\lambda_0 > 0$.

To determine then the performance characteristic of Method 1, i.e., the probability (1.1), it is clear that we are concerned with the conditional probability

$$\pi_1 = \frac{P\{|u| > t_\varepsilon, |z| \leqslant t_\varepsilon\}}{P\,\{|u| > t_\varepsilon\}}, \qquad \ldots \quad (2.5)$$

since $\lambda_0 > 0$ implies that $|u| > t_\varepsilon$. Now z is a normal variate with zero mean and unit variance, its correlation with u being

$$\rho_1 = (\rho-\alpha)/\sqrt{1-2\rho\alpha+\alpha^2}.$$

Since u and z have a bivariate normal distribution, the numerator in (2.5) is

$$\frac{1}{(2\pi)\sqrt{1-\rho_1^2}}\int\!\!\int_{R'} e^{-\frac{1}{2(1-\rho_1^2)}\{(u-\xi)^2 - 2\rho_1(u-\xi)z + z^2\}}\,du\,dz \qquad \ldots \quad (2.6)$$

the region R' in the (u, z) plane consisting of that part of the strip : $|z| \leqslant t_\varepsilon$ outside the lines : $|u| = t_\varepsilon$. The denominator in π_1 is available from tables of the normal integral

$$P\{|u| > t_\varepsilon\} = 1 - \frac{1}{\sqrt{2\pi}}\int_{-t_\varepsilon-\xi}^{t_\varepsilon-\xi} e^{-\frac{1}{2}\tau^2}\,d\tau \qquad \ldots \quad (2.7)$$

which approaches unity as $\xi \to \infty$.

Note also that the restriction that the discriminant Λ be positive is not involved in (2.5), since as Fieller (1944) has shown, if $|u| < t_\varepsilon$ the expression :

$$u^2-2\rho uv+v^2 > t_\varepsilon^2(1-\rho^2), \text{ or } \Lambda > 0.$$

We shall defer setting down the final expression of (2.5) in order to consider a second method or interpretation of the confidence limits for α presumed to be available from this quadratic method.

The following is believed to be the correct mathematical interpretation of the method given in Fisher, (1948) for determining confidence limits, or rather 'fiducial limits' for α on the basis of the inequality (2.2).

Method 2. *The equation* (2.2) *yields confidence limits* $\underline{\alpha}_1 = (\lambda_1-\sqrt{\Lambda})/\lambda_0$, $\bar{\alpha}_1 = (\lambda_1+\sqrt{\Lambda})/\lambda_0$ *provided* $\lambda_0 \geqslant 0$; *if* $\lambda_0 < 0$, *then the confidence limits are* : $\underline{\alpha}_2 = (\lambda_1+\sqrt{\Lambda})/\lambda_0$, $\bar{\alpha}_2 = (\lambda_1-\sqrt{\Lambda})/\lambda_0$.

For the performance characteristic π_2 then

$$\pi_2 = \frac{P_1+P_2}{P\{\Lambda > 0\}} \qquad \dots \quad (2.8)$$

in terms of the respective probabilities

$$P_1 = P\{|u| > t_\varepsilon,\ |z| < t_\varepsilon\} \qquad \dots \quad (2.9)$$

or the probability of obtaining the limits $\underline{\alpha}_1$, $\bar{\alpha}_1$, $(\lambda_0 > 0)$, and

$$\begin{aligned} P_2 &= P\{|u| < t_\varepsilon,\ |z| > t_\varepsilon,\ \Lambda > 0\} \\ &= P\{|u| < t_\varepsilon,\ |z| > t_\varepsilon,\ u^2-2\rho uv+v^2 > t_\varepsilon^2(1-\rho^2)\}, \end{aligned} \qquad \dots \quad (2.10)$$

with respect to the limits $\underline{\alpha}_2$, $\bar{\alpha}_2$, $(\lambda_0 < 0)$. It is important to note in P_2 that the inequality in z becomes $|z| > t_\varepsilon$, since in using the limits $\underline{\alpha}_2$, $\bar{\alpha}_2$ when the leading coefficient λ_0 of the parabola : $\lambda_0\alpha^2+2\lambda_1\alpha+\lambda_2$ in α is < 0, we are in fact considering the region: $|z| > t_\varepsilon$.

As regards the denominator in (2.8) this may be computed directly if we make the transformations:

$$\begin{aligned} t &= \frac{1}{\sqrt{2(1+\rho)}}(u+v) \\ w &= \frac{1}{\sqrt{2(1-\rho)}}(v-u) \end{aligned} \qquad \dots \quad (2.11)$$

with $E(t) = \xi(1+\alpha)/\sqrt{2(1+\rho)}$, $E(w) = \xi(\alpha-1)/\sqrt{2(1-\rho)}$

such that (t, w) are uncorrelated and have unit variance. Hence

$$P\{\Lambda > 0\} = P\{t^2 + w^2 > t_\varepsilon^2\} = \mathrm{P}\{\chi_2'^2 > t_\varepsilon^2\}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} P\{\chi_{2k+2}^2 > t_\varepsilon^2\} \qquad \ldots \quad (2.12)$$

in terms of the non-central chi-square distribution with two degrees of freedom. Here $\lambda = \xi^2(1-2\rho\alpha+\alpha^2)/2(1-\rho^2)$. The denominator (2.12) may be computed with the aid of tables of the Poisson distribution.

To deal again with the probabilities π_1, π_2, it is possible now to utilize a geometric method based on certain properties of the bivariate normal surface which enables us to compute these.

Consider then the numerator in (2.8). In the original (u, v) plane the parallel lines : $v = \alpha u \pm t_\varepsilon \sqrt{1-2\rho\alpha+\alpha^2}$ with slope equal to α are tangent to the ellipse: $u^2 - 2\rho uv + v^2 = t_\varepsilon^2(1-\rho^2)$. Consider then the region of integration as the transformed (u, z) plane.

Thus the region of integration originally indicated in Figure 1a becomes the shaded portion in Figure 1b.

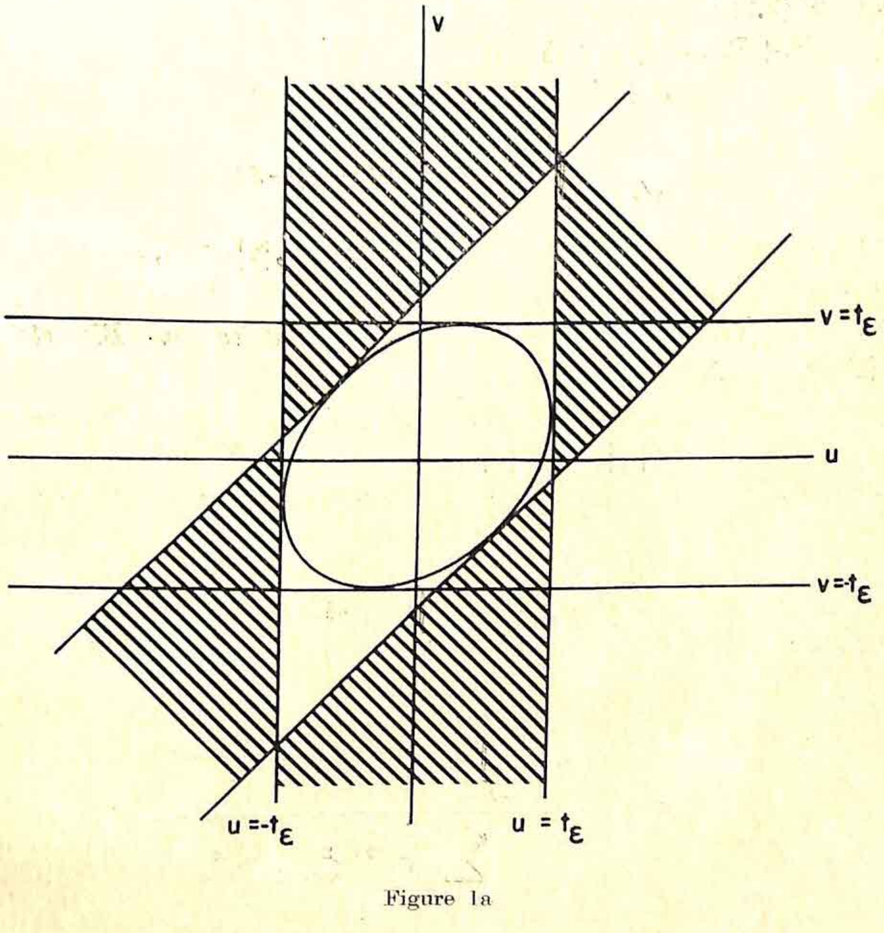


Figure 1a

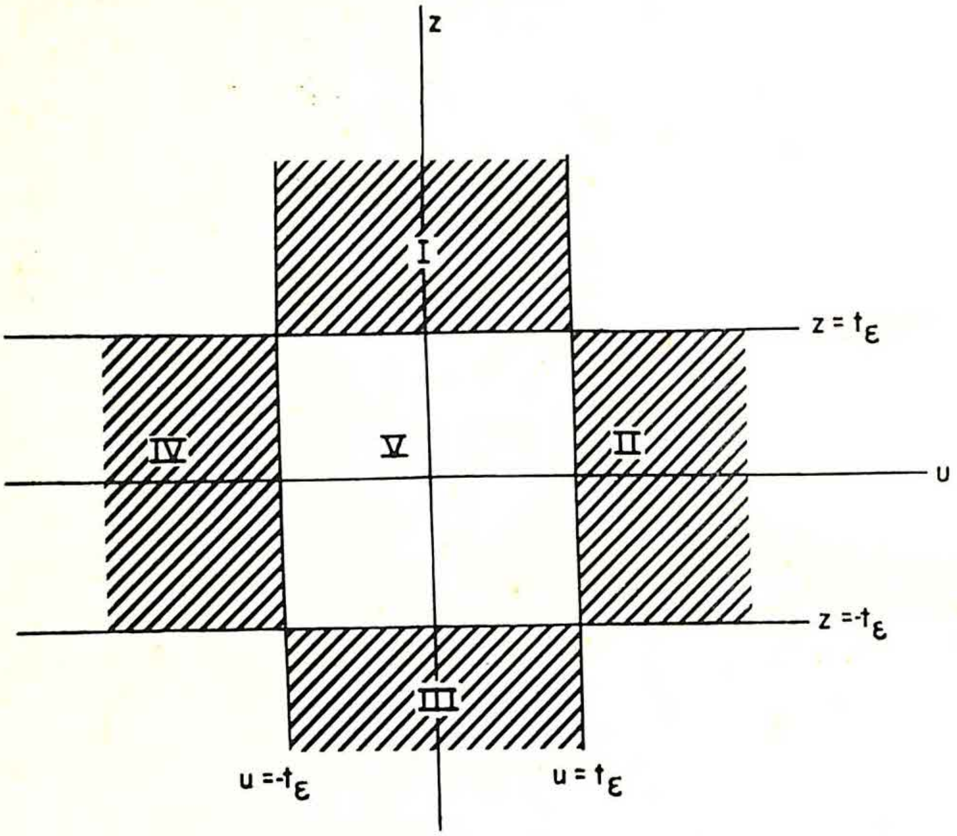


Figure 1b

The numerator of (2.8) becomes

$$P_1+P_2 = \text{I}+\text{II}+\text{III}+\text{IV}$$

$$= (\text{I}+\text{III}+\text{V})+(\text{II}+\text{IV}+\text{V})-2\text{V}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-t_\varepsilon-\xi}^{t_\varepsilon-\xi} e^{-\frac{1}{2}\tau^2}\,d\tau+(1-\epsilon)-2V \qquad \ldots \quad (2.13)$$

where V is the bivariate normal integral, i.e., $P\{|u|< t_\varepsilon, |z|< t_\varepsilon\}$, or the volume of the bivariate normal surface in (u, z) with the rectangular base: $-\xi-t_\varepsilon < u < -\xi+t_\varepsilon$, $-t_\varepsilon < z < t_\varepsilon$. This is available from the *Tables for Biometricians and Statisticians* (Pearson, 1931).

As regards the numerator of π_1, it is easily seen from Figure 1b that

$$P\{|u|> t_\varepsilon,\ |z|< t_\varepsilon\} = (1-\epsilon)-V. \qquad \ldots \quad (2.14)$$

The final results for the performance characteristics for the two methods are then

$$\pi_1 = \frac{(1-\epsilon)-V}{1-\frac{1}{\sqrt{2\pi}}\int_{-t_\varepsilon-\xi}^{t_\varepsilon-\xi} e^{-\frac{1}{2}\tau^2}\,d\tau}$$

$$\pi_2 = \frac{\frac{1}{\sqrt{2\pi}}\int_{-t_\varepsilon-\xi}^{t_\varepsilon-\xi} e^{-\frac{1}{2}\tau^2}\,d\tau+(1-\epsilon)-2V}{e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}P\{\chi^2_{2u+2} > t_\varepsilon^2\}}, \qquad \ldots \quad (2.15)$$

as functions of α, ξ, ρ. The performance characteristics π_1, π_2 have been computed for selected values in Tables 1–2 and $t_\varepsilon = 1.96$, i.e., with reference to an assumed confidence coefficient = 0.95.

TABLE 1. PERFORMANCE CHARACTERISTIC $\pi_1(\alpha, \xi, \rho)$ (METHOD 1).

	$\xi=0.25$		$\xi=0.50$		$\xi=0.75$		$\xi=1.0$		$\xi=1.5$	
α \ ρ	0	0.5	0	0.5	0	0.5	0	0.5	0	0.5
0.5	.849	.950	.873	.950	.896	.950	.914	.950	.938	.950
1.0	.694	.829	.764	.859	.827	.887	.872	.908	.926	.938
2.5	.414	.367	.621	.597	.762	.755	.853	.843	.922	.921
5.0	.348	.348	.596	.596	.755	.755	.843	.843	.921	.921

TABLE 2. PERFORMANCE CHARACTERISTIC $\pi_2(\alpha, \xi, \rho)$ (METHOD 2)

	$\xi=0.25$		$\xi=0.50$		$\xi=0.75$		$\xi=1.0$		$\xi=1.5$	
α \ ρ	0	0.5	0	0.5	0	0.5	0	0.5	0	0.5
0.5	.575	.654	.576	.667	.586	.689	.604	.714	.661	.773
1.0	.442	.554	.426	.553	.419	.560	.429	.576	.491	.636
2.5	.206	.172	.164	.182	.177	.189	.205	.212	.327	.330
5.0	.089	.086	.082	.080	.112	.111	.167	.167	.323	.322

Thus if $\alpha = 2.5$, $\xi = 0.5$, $\rho = 0$, the probability of being correct in determining the presumed 95% confidence limits by Method 1 is only 0.621, while by Method 2 the probability is 0.164.

Regarding the limiting values of the performance characteristic $\pi_1(\alpha, \xi, \rho)$ of Method 1, it is easy to verify that

$$\lim_{\alpha \to 0} \pi_1(\alpha, \xi, \rho) = 1-\varepsilon. \qquad \ldots \quad (2.16)$$

As regards the limit as $\alpha \to +\infty$, it also may be shown that

$$\lim_{\alpha \to \infty} \pi_1(\alpha, \xi, \rho) = \frac{(1-\varepsilon) - \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-t_\varepsilon-\xi}^{t_\varepsilon-\xi} e^{-\frac{1}{2}\tau^2}\, d\tau}{1 - \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-t_\varepsilon-\xi}^{t_\varepsilon-\xi} e^{-\frac{1}{2}\tau^2}\, d\tau}, \qquad \ldots \quad (2.17)$$

which may be further evaluated accordingly as $\xi \gtrless 2t_\varepsilon$, etc. Finally

$$\lim_{\xi \to \infty} \pi_1(\alpha, \xi, \rho) = 1-\varepsilon.$$

3. CONFIDENCE LIMITS FOR AN ABSCISSA

We now consider the appropriate confidence limits in the following regression problem. Suppose that $\{y_i\}$, $(i = 1, \ldots, n)$, represent n observations on a normally distributed variate y, each observation being obtained at a separate though fixed level or value x_i, $(i = 1, \ldots, n)$. It is further assumed that the regression of y on x is linear, i.e., $E(y_i) = \alpha+\beta x_i$, $(i = 1, \ldots, n)$, α, β being certain unspecified parameters and that each y has a common variance $= \sigma^2$ (known). Corresponding to a particular value $= Y$ of the variate y it is required to determine confidence limits for the value of the abscissa X. This problem is a familiar one in regression analysis and of particular interest in biological assay and standardization problems. (Bliss, 1935 ; Irwin, 1943).

Formally then we are considering an appropriate test of the hypothesis

$$Y = \alpha+\beta X \qquad \ldots \quad (3.1)$$

with α, β unspecified and X, Y certain values.

Utilizing the familiar properties of orthogonal transformations, the problem may be reduced to considering a sequence of n normally and independently distributed variates $\{w_i\}$, $(i = 1, \ldots, n)$, with variance $= \sigma^2$,

such that
$$E\,(w_k) = 0, \quad (k = 1, \ldots, n-2),$$
$$E\,(w_{n-1}) = \theta_{n-1} = \sqrt{\Sigma_{n-1}}\,\beta,$$
$$E\,(w_n) = \theta_n = \sqrt{n}\alpha+u_n\beta,$$

the u's being fixed values, $\qquad \Sigma_{n-1} = \sum_{i=1}^{n-1} u_i^2 = \sum_{i=1}^{n} (x_i-\bar{x})^2$,

and in particular $\qquad \sum_{i=1}^{n-2} w_i^2 = \sum_{i=1}^{n} \{y_i-\bar{y}-b(x_i-\bar{x})\}^2$.

The hypothesis (3.1) is equivalent to testing : $\theta_n = A\theta_{n-1}+B$, the coefficients of this linear hypothesis being $A = (u_n - \sqrt{n}\,X)/\sqrt{\Sigma_{n-1}}$, $B = \sqrt{n}\,Y$. The appropriate region for testing this hypothesis is

$$|z^*| = \frac{|w_n-Aw_{n-1}-B|}{\sigma\sqrt{1+A^2}} > t_\varepsilon, \qquad \ldots \quad (3.2)$$

t_ε being specified from tables of the normal distribution with level of significance $= \varepsilon$. Hence confidence limits for X are then available from

$$|z^*| = \frac{|\,Y-\bar{y}-b\,(X-\bar{x})}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{(X-\bar{x})^2}{\Sigma(x-x)^2}}} \leqslant t_\varepsilon \qquad \ldots \quad (3.3)$$

in terms of the original variates, $\bar{x}$, $\bar{y}$ being the means of x and y respectively, and b the sample regression coefficient of y on x.

Instead of computing confidence limits for X, it is more convenient to consider the corresponding limits for $X^* = A = (u_n - \sqrt{n}\,X)/\sqrt{\Sigma_{n-1}}$, a fixed linear function of X. The confidence limits for X^* are then obtained on the basis of the inequality

$$\lambda_0^* X^{*2} - 2\lambda_1^* X^* + \lambda_2 \leqslant 0 \qquad \ldots \quad (3.4)$$

with

$$\lambda_0^* = w_{n-1}^2 - \sigma^2 t_\varepsilon^2$$

$$\lambda_1^* = w_{n-1}(w_n - \sqrt{n}\,Y)$$

$$\lambda_2^* = (w_n - \sqrt{n}\,Y)^2 - \sigma^2 t_\varepsilon^2$$

and discriminant

$$\Lambda^* = \sigma^2 t_\varepsilon^2 \,[w_{n-1}^2 + (w_n - \sqrt{n}\,Y)^2 - \sigma^2 t_\varepsilon^2\,].$$

Since Methods 1 and 2 are extensively used in computing confidence limits for X it is of interest to ascertain their performance characteristics.

The performance characteristic for Method 1 is

$$\pi_1^*(X, \beta, \sigma) = \frac{P\,\{|z^*| < t_\varepsilon, |w_{n-1}| > \sigma t_\varepsilon\}}{P\,\{|w_{n-1}| > \sigma t_\varepsilon\}}. \qquad \ldots \quad (3.5)$$

It will be noted that the geometric situation is just as indicated in Figures 1a-1b. Instead of the ellipse there is a circle : $w_{n-1}^2 + (w_n - \sqrt{n}\ Y)^2 = \sigma^2 t_\varepsilon^2$ with center at $(0, \sqrt{n}\ Y)$ in the (w_{n-1}, w_n) plane. We then transform to the bivariate normal space of the (w_{n-1}, z^*) variates, which have mean values $\sqrt{\Sigma_{n-1}}\ \beta$, 0 respectively, variances σ^2, 1 and correlation $= -A/\sqrt{1+A^2} = -X^*/\sqrt{1+X^{*2}}$. If we denote by V^* the volume of the normal surface with rectangular base :

$$-t_\varepsilon - \frac{1}{\sigma}\sqrt{\Sigma_{n-1}}\,\beta \leqslant w_{n-1} \leqslant t_\varepsilon - \frac{1}{\sigma}\sqrt{\Sigma_{n-1}}\,\beta,\ -t_\varepsilon \leqslant z^* \leqslant t_\varepsilon,$$

then the performance characteristics for Methods 1 and 2 are

$$\pi_1^*(X, \beta, \sigma) = \frac{(1-\epsilon) - V^*}{1 - \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-t_\varepsilon - \frac{1}{\sigma}\sqrt{\Sigma_{n-1}}\beta}^{t_\varepsilon - \frac{1}{\sigma}\sqrt{\Sigma_{n-1}}\beta} e^{-\frac{1}{2}\tau^2}\, d\tau} \qquad \ldots \quad (3.6)$$

$$\pi_2^*\,(X, \beta, \sigma) = \frac{\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-t_\varepsilon - \frac{1}{\sigma}\sqrt{\Sigma_{n-1}}\beta}^{t_\varepsilon - \frac{1}{\sigma}\sqrt{\Sigma_{n-1}}\beta} e^{-\frac{1}{2}\tau^2} d\tau + (1-\epsilon) - 2V^*}{e^{-\lambda^*} \displaystyle\sum_{k=0}^{\infty} \frac{\lambda^{*k}}{k\,!}\, P\,\{\chi_{2k+2}^2 > t_\varepsilon^2\}} \qquad \ldots \quad (3.7)$$

with

$$\lambda^* = (\Sigma_{n-1})(1+X^{*2})\beta^2/2\sigma^2.$$

Values of $\pi_1^*(X, \beta, \sigma)$, $\pi_2^*(X, \beta, \sigma)$ may then be obtained from Tables 1 and 2 with X^* replacing α, etc.

Bureau Ednl. Psy. Research
TRAINING COLLEGE
Dated

4. CONFIDENCE LIMITS FOR INTERSECTION POINT OF TWO LINES

As a further example of the quadratic methods for determining confidence limits consider the following problem.

Suppose that (x_{1i}, y_{1i}), $(i = 1, \ldots, n_1)$, (x_{2j}, y_{2j}), $(j = 1, \ldots, n_2)$, represent two independent samples of n_1 and n_2 pairs of observations respectively. It is assumed that for fixed values of the x's the y's are normally distributed with means

$$\begin{aligned} E(y_1) &= \alpha_1+\beta_1 x_1 \\ E(y_2) &= \alpha_2+\beta_2 x_2 \end{aligned} \qquad \ldots \quad (4.1)$$

and with a common fixed variance $=\sigma^2$. On the basis of the two samples it is required to determine confidence limits for the abscissa X of the intersection point of the two population regression lines (Fisher, 1948). As a problem in testing hypotheses it is seen that it is required to find an appropriate region for the hypothesis : $\alpha_1+\beta_1 X = \alpha_2+\beta_2 X$. We shall however not consider this problem as such but rather in an equivalent and simpler formulation.

The problem of determining confidence limits for the intersection point of two regression lines is equivalent to that of determining confidence limits for the intersection point of one regression line with a certain *fixed* line in the plane. For convenience we may assume that the fixed line happens to go through the origin, and then concern ourselves merely with one sample of n pairs of observations (x_i, y_i), $(i = 1, \ldots, n)$. For each fixed x the regression $E(y) = \alpha+\beta x$ is linear with a constant known variance $=\sigma^2$ and further there exists a fixed line in the plane: $y = \beta_0 x$ (β_0 known). It is then required to test the hypothesis: $\beta_0 X = \alpha+\beta X$.

The canonical reduction of the problem is the same as that in section 3, so that we may merely consider the sequence $\{w_i\}$, $(i = 1, \ldots, n)$, such that $E(w_k) = 0$, $(k = 1, \ldots, n-2)$, $E(w_{n-1}) = \theta_{n-1} = \sqrt{\Sigma_{n-1}}\beta$, $E(w_n) = \theta_n = \sqrt{n}\,\alpha+u_n\beta$. The hypothesis merely states that : $\theta_n = A\theta_{n-1}+C$, where $C = \sqrt{n}\beta_0 X$. Confidence limits are then available from

$$|z^{**}| = \frac{|w_n-A\,w_{n-1}-C|}{\sigma\sqrt{1+A^2}}$$

$$= \frac{|(w_n-\beta_0 u_n)-A(w_{n-1}-\sqrt{\Sigma_{n-1}}\beta_0)|}{\sigma\sqrt{1+A^2}} \leqslant t_\varepsilon. \qquad \ldots \quad (4.2)$$

Since A is a linear function of X, confidence limits for $A(= X^{**})$ are then available from the inequality,

$$\lambda_0^{**}\, X^{**2} - 2\lambda_1^{**}\, X^{**} + \lambda_2^{**} \leqslant 0 \qquad \ldots \quad (4.3)$$

where

$$\lambda_0^{**} = (w_{n-1}-\beta_0\sqrt{\Sigma_{n-1}})^2-\sigma^2 t_\varepsilon^2$$

$$\lambda_1^{**} = (w_{n-1}-\beta_0\sqrt{\Sigma_{n-1}})(w_n-\beta_0 u_n)$$

$$\lambda_2^{**} = (w_n-\beta_0 u_n)^2-\sigma^2 t_\varepsilon^2$$

and

$$\Lambda^{**} = (\sigma^2 t_\varepsilon^2)[(w_{n-1}-\beta_0\sqrt{\Sigma_{n-1}})^2+(w_n-\beta_0 u_n)^2-\sigma^2 t_\varepsilon^2].$$

Hence
$$\pi_1^{**}(X,\alpha,\beta,\beta_0) = \frac{P\{|z^{**}| < t_\varepsilon, |w_{n-1}-\beta_0\sqrt{\Sigma_{n-1}}| > t_\varepsilon\sigma\}}{P\{|w_{n-1}-\beta_0\sqrt{\Sigma_{n-1}}| > t_\varepsilon\sigma\}}$$
$$= \frac{(1-\epsilon)-V^{**}}{1-\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-t_\varepsilon-\frac{1}{\sigma}(\beta-\beta_0)\sqrt{\Sigma_{n-1}}}^{t_\varepsilon-\frac{1}{\sigma}(\beta-\beta_0)\sqrt{\Sigma_{n-1}}} e^{-\frac{1}{2}\tau^2}\,d\tau} \qquad \dots \quad (4.4)$$

and

$$\pi_2^{**}(X,\alpha,\beta,\beta_0) = \frac{\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-t_\varepsilon-\frac{1}{\sigma}(\beta-\beta_0)\sqrt{\Sigma_{n-1}}}^{t_\varepsilon-\frac{1}{\sigma}(\beta-\beta_0)\sqrt{\Sigma_{n-1}}} e^{-\frac{1}{2}\tau^2}\,d\tau+(1-\epsilon)-2V^{**}}{e^{-\lambda^{**}}\displaystyle\sum_{k=0}^{\infty}\frac{\lambda^{**k}}{k!}P\{\chi^2_{2k+2} > t_\varepsilon^2\}}$$

in terms of $\lambda^{**} = \frac{1}{2\sigma^2}[\{\sqrt{n}\alpha+(\beta-\beta_0)u_n\}^2+(\beta-\beta_0)^2\Sigma_{n-1}]$, V^{**} being the volume under the bivariate normal surface in (z^{**}, w_{n-1}) in the region : $-t_\varepsilon-\frac{1}{\sigma}\sqrt{\Sigma_{n-1}}\beta_0 < w_{n-1} < t_\varepsilon+\frac{1}{\sigma}\sqrt{\Sigma_{n-1}}\beta_0$, $-t_\varepsilon < z^{**} < t_\varepsilon$. The correlation between z^{**} and w_{n-1} is $= -A/\sqrt{1+A^2} = -\sqrt{\Sigma_{n-1}}(u_n-\sqrt{n}X)/\sqrt{\Sigma_{n-1}+(u_n-\sqrt{n}X)^2}$ as a function of X. Tables 1 and 2 may then be utilized in evaluating π_1^{**}, π_2^{**}.

The performance characteristics π_1, π_2 for Methods 1 and 2 have also been computed for the three situations described above when now the population variances are unknown. Clearly the same situations obtain as indicated by the limiting cases described above, in that π_1, π_2 are functions of the particular parameters for which confidence limits are supposed to be provided.

NOTE AND ACKNOWLEDGEMENTS

The preceding results were announced some years ago (Abstracts. *Ann. Math. Stat.* vol. 19, 1948) but their publication was unfortunately delayed. In view of the continuing interest in the problem of the interval estimation of a ratio, etc. (Fieller, 1954) it was decided to present some of the results at this time in this article.

The author wishes to acknowledge his indebtedness to Professor W. G. Cochran, now of the School of Hygiene and Public Health of John Hopkins University, in first suggesting the problem, and in particular to Professor J. Neyman of the University of California, Berkeley for his assistance in proposing and developing specific methods and solutions for the performance characteristics in section 2. Thanks are also due to Dr. Evelyn Fix for checking some of the Tables presented in section 2.

REFERENCES

BLISS, C. I. (1935) : The comparison of dosage mortality data. *Ann. Applied Biology*, **22**, 307–333.

FIELLER, E. C. (1940) : Biological standardization of insulin. *Suppl. J. Roy. Stat. Soc.*, **7**, 1–64.

———, (1944) : A fundamental formula in the statistics of bioassay. *Quart. J. Pharm. and Pharmacology*, **17**, 117–123.

———, (1954) : Symposium on interval estimation. *J. Roy. Stat. Soc.*, B, **16**, 175–222.

FISHER, R. A. (1948) : *Statistical Methods for Research Workers*, section 16.2, 10th or subsequent editions, Oliver and Boyd.

IRWIN, J. (1943) : On the calculation of the error of biological assays. *J. Hygiene*, **43**, 121–128.

NEYMAN, J. (1937) : Outline of theory of statistical estimation based on the classical theory of probability. *Phil. Trans. Roy. Soc.* A, **236**, 333–380, *London*.

PEARSON, K. (1931) : *Tables for Biometricians and Statisticians*, **2**, Cambridge University Press.

Paper received : June, 1956.

SENSITIVITY OF A PROPOSED METHOD OF QUALITY CONTROL

By W. L. STEVENS
Universidade de São Paulo, Brazil

In a recent paper, B. C. Bhattacharyya (1955) makes the interesting suggestion that when an operation is controlled by "go" and "no go" gauges, the sensitivity of control may be enhanced by making use of the exact measurements of articles which fall outside the control limits. In 1948, I proposed (the proposal was not new) the use of the indicator, $c-a$, to show when the mean goes out of control, where c and a are the numbers of articles in a sample of n, which respectively exceed the upper gauge or fall short of the lower gauge. Bhattacharyya's proposal is that, instead of giving equal weights to "rejected" articles, they shall receive weights proportional to the absolute amounts by which the measurements differ from the gauge which rejected them. A similar modification is made to the indicator, $a+c$, used for detecting a change in standard deviation. He further says :

> "It is likely that X [the indicator proposed], based on actual measurements, will yield better results than Stevens' indicator $c-a$, but a comparison of the 'powers' of the two will not be undertaken."

As the proposed new method demands additional work at the inspection bench and, moreover, work which is more difficult (since measurement needs more skill than gauging), it does seem to the present writer that the question of what is likely to be gained is a very important one. We therefore propose to examine in this paper the sensitivity of the new method for detecting changes in the mean.

We shall suppose that the population is normal and, without essential loss of generality, that the mean is zero and the standard deviation unity. We suppose further the "go" and "no go" gauges have magnitudes $-x$ and $+x$. The proposed method is that each article receives a score,

$$\begin{array}{ll} t-x & \text{if } t > x \\ 0 & \text{if } -x < t < x \\ t+x & \text{if } t < -x \end{array}$$

where t is the measurement (made only if the article is 'rejected'). The indicator, X, is then the sum of the scores in a sample of n.

The expected value of the score is found as the sum of the contributions from two classes: those articles which exceed the larger gauge and those which fall short of the lower gauge. From the former class the expectation is evidently

$$\int_x^\infty (t-x)\, z(t)dt$$

where
$$z(t) = \exp\,(-t^2/2)/\sqrt{(2\pi)}.$$

The integral is readily found to be :

$$z-xp$$

where
$$z = z(x)$$

and
$$p = \int_x^\infty z(t)dt.$$

The contribution from the second class ($t < -x$) is evidently the negative of the first, so that the expected score is zero.

We now want the expected score when the mean shifts a small distance $\delta\mu$ away from zero. The contribution of the first class may be found by recognizing that a movement of the mean through $\delta\mu$ to the right is equivalent to moving the upper gauge $\delta\mu$ to the left. In other words, we substitute $x-\delta\mu$ for x in the above formula, obtaining

$$(z-xp)-\{d\,(z-xp)/dx\}\,\delta\mu$$

$$= (z-xp)-(-xz-p+xz)\delta\mu$$

$$= (z-xp)+p\delta\mu.$$

The contribution to the expected score is therefore increased by $p\delta\mu$. Clearly, the increase in the contribution from the second class is the same, and we conclude that the expected score is $2p\delta\mu$. The estimate of $\delta\mu$ is therefore simply

$$\delta m = \frac{X}{2np}.$$

The variance of a single score is

$$2\int_{x}^{\infty}(t-x)^2\,z(t)\,dt = 2\left\{\int t^2 z\,(t)dt - 2x\int tz\,(t)\,dt + x^2\int\ z\,(t)\,dt\right\}$$

$$= 2\{(xz+p)-2xz+x^2p\}$$

$$= 2\{(1+x^2)p-xz\}.$$

The variance of X is n times this, and the variance of the estimate δm is therefore,

$$\frac{(1+x^2)p-xz}{2np^2}.$$

As the variance of the mean of n measurements is $1/n$, the efficiency of estimation of the mean in the region of zero is accordingly :

$$\frac{2p^2}{(1+x^2)p-xz}.$$

The efficiencies percentage of the indicator $c-a$ and of the new indicator are tabulated below, the values for the former being copied from Stevens (1948).

EFFICIENCIES (PERCENTAGE)

p	$c-a$	new	m.l.	$E\frac{(2+p)}{(2-p)}$
0.5	63.7	100	100	106.2
0.4	74.6	97.6	99.6	111.9
0.3	80.6	89.9	96.5	109.0
0.25	80.8	83.7	92.9	103.9
0.2	78.4	75.5	87.1	95.8
0.15	72.5	64.8	78.3	84.3
0.1	61.6	50.9	62.4	68.1
0.05	42.6	32.0	43.9	44.8
0.01	14.2	9.4	14.4	14.5

It appears that, although for large p, the efficiencies of the new method are indeed higher, at some point between $p = 0.25$ and 0.2 they are equal and for smaller values of p the efficiencies of the new method are below those of the old method. A higher efficiency for the new method can therefore be obtained only if we are prepared to do much more work of measurement. Bhattacharyya proposes using gauges at the quartiles, implying that half the articles inspected will have to be measured. What would be obtained for this great increase in labour, is an increase of efficiency from 80.8 to 83.7 per cent, an improvement which could equally be obtained by doing no measurement but increasing the number of gaugings by one twenty-eighth !

It is of interest to see how the measurements of articles falling outside the gauge limits could be used to the maximum advantage. The logarithm of the likelihood is

$$-\Sigma(t-\delta\mu)^2/2+b\ \log\ (q+\delta q)$$

where b is the number of articles falling between the gauges and Σ is over articles falling outside the gauges, and

$$q+\delta q = \int_{-x}^{+x} z(t-\delta\mu)dt$$

$$= \int_{-x}^{+x} \{z+tz\delta\mu+0\,(\delta\mu^2)\}\ dt$$

$$= q+0(\delta\mu^2).$$

It follows that, for small displacements of the mean, the second term does not contribute anything and therefore that $T = \Sigma t$ contains all the information. The individual score is accordingly t, which represents the measurement if this lies outside the gauges and zero if within.

The expected value of t is obtained as the sum of contributions from the two classes; above the upper gauge and below the lower gauge respectively. From the former we have

$$\int_{x}^{\infty} tz\,(t-\delta\mu)dt = \int_{x}^{\infty} \{tz\,(t)+t^2z\,(t)\delta\mu\}dt = z+(xz+p)\delta\mu.$$

The other contribution is clearly $-z+(xz+p)\delta\mu$,

so that the expected score is $2(xz+p)\delta\mu$

and the estimate of $\delta\mu$ is $\delta m = \dfrac{T}{2n(xz+p)}$.

The variance of the individual score, when the mean is zero, is

$$2\int_{x}^{\infty} t^2 z(t)\, dt = 2(xz+p).$$

The variance of the estimate δm is hence

$$\frac{2n(xz+p)}{4n^2(xz+p)^2} = \frac{1}{2n(xz+p)}$$

and the efficiency is $2(xz+p)$. The values of the efficiency are shown in the penultimate column of the table, headed 'm.l.' The efficiency of the proposed method falls appreciably short of that of the most efficient method.

Now let us make an assumption which is the most favourable possible for the use of measurements on 'rejected' articles—that the cost of a measurement is the same as the cost of a gauging (actually it will be greater). Then, if measurements are used, in each sample of n we shall make, on the average, $n(2-p)$ gaugings and $2np$ measurements, i.e., $n(2+p)$ observations. Alternatively, if costs are proportional to number of observations we could use a sample of $n(2+p)/(2-p)$ articles and only gauge them. The quantity which measures the information supplied by the second method is therefore obtained by multiplying E, the efficiency of simple gauging (column, $c-a$, of the table) by $(2+p)/(2-p)$. The results are shown in the last column of the table.

It is observed that, for equal number of observations, the method of gauging alone, using the indicator $c-a$, gives a more sensitive control than can be obtained by using measurements and the best indicator, I. (Actually, the superiority will be lost for very small values of p, outside the table, but that is of no practical consequence because such small values would not be used in either method). Since we have assumed conditions unduly favourable to the proposal to use measurements on 'rejected' articles, it does appear that the proposal is not likely to prove very fruitful.

Admittedly, the investigation is in the realm of 'large samples' and, to complete the study, one should also compare the whole course of the power curves in small samples. In view of the results of section 5.2 of Stevens, (1948), however, we are confident that a fuller investigation would not produce substantially different results.

CONCLUSION

The most costly but most sensitive method of control is by exact measurement of every sampled article. The cheapest and simplest method is by the use of two gauges. One asks if, between these extremes, there exists a method which, in a suitable combination of costs, may be shown to be superior to either of the extremes. Bhattacharyya's suggestion for such a method is first to submit the articles to a pair of gauges and afterwards to measure all which are 'rejected' by the gauges. The examination of this method has shown, however, that it will not usually be economical.

Seeking for another method, we may move 'inwards' from either extremes. Thus, instead of making 'exact' measurements, we may use an extremely coarse scale. Such a method, especially if the divisions are at the half intervals (0.5, 1.5, 2.5, etc.) is excellent in practice, because it is less subject to mistakes and avoids the frequently observed bias towards the more strongly marked division lines of the conventional scale. The loss of sensitivity is small; even with a scale interval of one standard deviation, the efficiency of estimation of the mean is 91.7% and of the standard deviation, 83.3%.

Alternatively, we may consider the possibility of using four gauges instead of two. It can be shown for example, that if the gauges are at $-1.8, -0.6, +0.6$ and $+1.8$, the efficiency for the mean is 90.0% while that for the standard deviation is 75.3%. Although four gauges are used, the average number of gaugings per article is only 2.27.

These two possibilities — the use of measurements on a very coarse scale, and the use of four gauges — seem to cover, in practice, the range between the two extremes of exact measurement and control by two gauges and it appears unlikely that yet another method can be found which, with any reasonable assumptions of the costs of the various operations, can be shown to be superior to all the others.

References

Bhattacharyya, B. C. (1955): A possible use of certain measurements in control by gauging. *Sankhyā*, **15**, 210-213

Stevens, W. L. (1948): Control by gauging. *J. Roy. Stat. Soc.*, **10**, 54-108.

Paper received : June 1956.

THE NATIONAL SAMPLE SURVEY

NUMBER 8

REPORT ON
PRELIMINARY SURVEY OF URBAN UNEMPLOYMENT
SEPTEMBER 1953

FOREWORD

0.1. In India a large number of persons are without employment or are obliged to remain idle a part of the time for lack of gainful work. The population is growing at the rate of about 40 or 45 lakhs per year and about 18 or 20 lakhs of persons normally start seeking work for the first time every year. Unemployment and underemployment constitute the greatest social problem in the country at present. This is why it has been decided that an important aim of planning in India must be to get rid of unemployment in 10 years. There must not be any increase in unemployment during the Second Five Year Plan, and efforts must be made to create enough work to absorb at least the new entrants into the labour force during this period. An adequate understanding of the problem is a prerequisite for devising suitable means to deal with it. Unfortunately, very little is known about the precise extent or character of unemployment and underemployment in India. There have been occasional surveys in particular areas but the results are often not comparable; and it is not possible to make any estimates for the country as a whole.

0.2. Unemployment is not homogeneous. A single aggregate figure purporting to give the total number of unemployed persons has no doubt some value but is too abstract to supply a basis for effective action. The problem is complex and has many aspects. As regards the persons seeking work we have to make a distinction between the two sexes, between different age-groups, between different educational and technical qualifications. We have also to take note of previous experience, community to which the person belongs, his mother tongue, whether he is a displaced person or an immigrant from another part of the country, duration of unemployment and many such personal matters. These special needs and characteristics of each individual who is working or who is seeking work have to be viewed in relation to the avenues of employment. The creation of opportunities of employment cannot thus be approached merely in overall terms. The problem has to be analysed in terms of sectors, regions and specific category of persons.

0.3. In India a very large number of persons work on their own account in household enterprises and do not hold any paid jobs and cannot therefore lose their jobs. The concept of unemployment, as usually adopted in the highly industrialised countries of the West, does not wholly apply to them. It is not true in their case that *during any particular reference period* of a week or a month, they are either fully employed or are wholly unemployed (which is the case in industrialised countries like the U.S.A.). Often they are fully occupied for part of the time and only partially or not at all for the rest of the period. It is the existence of this underemployment to an appreciable extent that makes the task of definition of unemployment and its measurement difficult. It becomes necessary to introduce the concept of the volume of work and earnings. That is to say, underemployment may be assessed with reference to the actual work done as a fraction of a "normal period of work", or the actual output as a fraction of the "normal output" or of actual earning as a fraction of the "normal earning". It may be necessary to construct reasonable standards of "normal period of work","normal output" and "normal earnings" and to assess the volume of employment on the basis of fractions of such normals or standards. The other important consideration is that the concept adopted should be suitable to enable quantitative evaluation being made. In fact, an important task before us is to formulate precisely the concepts and definitions of employment and unemployment suited to our requirements and to devise appropriate methods to measure them.

0.4. Unemployment and underemployment arise due to lack of balance between individual needs and capacity for work of the millions of persons who are seeking work and availabilities of paid jobs or of opportunities to undertake gainful work on own account. In industrially advanced countries there is full employment at full production. Unemployment can occur only when plants and factories are working below capacity, and fringe adjustments can be made through measures of social security. In India both unemployment and underemployment are chronic and structural because of a lack of economic development. Minor adjustments can, no doubt (and should, of course), be made through Employment Exchanges and similar devices but the fundamental remedy lies only in economic development.

0.5. It is necessary, therefore, to take a dynamic view of the way in which more and more paid posts would be created or greater opportunities would be opened up to facilitate earning on own account. It is also necessary to determine how educational and training facilities would be expanded to ensure an adequate supply of technical personnel which would be required in increasing number for economic development, and how housing, transport, medical welfare and other amenities would be provided in an increasing measure in future. Getting rid of unemployment and underemployment would then appear as the central task of planning from the point of view of both individual and national needs.

0.6. Surveys of unemployment and underemployment are thus of great value in connection with planning in India. The Planning Commission have naturally shown much interest in such surveys, and it has been decided to include employment surveys as a part of the programme of the National Sample Survey in future.

0.7. The present report gives an account of a preliminary survey which was undertaken in September 1953 covering towns with population above 50,000 but excluding the four big cities, Calcutta, Bombay, Madras and Delhi. This was the first survey of its kind and was undertaken as a pilot enquiry to help formulating suitable concepts and definitions and developing suitable designs and methods of surveys of employment, unemployment and underemployment in future. Although preliminary in character, the present survey has given for the first time a picture of urban unemployment in the country in quantitative terms.

0.8. A summary of the findings is given in Chapter Two of the Report. It may be briefly mentioned that in the urban areas covered in the present sample about 2.6 per cent of the population was unemployed in September 1953. If the rate of incidence of unemployment is about the same in smaller towns (with population less than 50,000) and if certain adjustments are made for the big cities then one might make an estimate of about 2.24 million unemployed out of an urban population of about 68 million in September 1953. It is estimated that about 2.74 million were severely underemployed, that is, had employment on own account for only a quarter or less of the normal period of work. Adding this 2.74 million, who also were on the verge of unemployment, the total urban unemployment would seem to be something like 5 million. In the absence of unemployment benefits, the urgency of the problem can be easily appreciated.

March 1956 P. C. MAHALANOBIS

J - 392
7.9.70

THE NATIONAL SAMPLE SURVEY

NUMBER 8

REPORT ON
PRELIMINARY SURVEY OF URBAN UNEMPLOYMENT
SEPTEMBER 1953

CONTENTS

		PAGE
FOREWORD	By P. C. Mahalanobis	19
CHAPTER ONE	Scope and Method	25
CHAPTER TWO	Summary of Results	30
CHAPTER THREE	General Characteristics of the Population ...	36
CHAPTER FOUR	The Employed and the Unemployed	46
STATISTICAL TABLES	Classified List of Tables	71
	Table (1)—(33)	74
APPENDIX 1	Comparison of Two Independent Sub-sample Estimates	122
APPENDIX 2	Schedule of Investigation	125

INDEX TO TABLES IN THE TEXT

CHAPTER ONE

PAGE

TABLE 1 Selection of sample towns 26

CHAPTER THREE

TABLE 2 : Percentage distribution of population by sex and age-group 36
TABLE 3 : Comparison of age distribution, Sample Survey and 1951 Census 37
TABLE 4 : Children, persons of working age and older persons in the population 37
TABLE 5 : Percentage distribution of males and females by economic status 38
TABLE 6 : Percentage distribution of population, males and females separately, by age-group and economic status 38
TABLE 7 : Percentage distribution of population by monthly household expenditure class .. 39
TABLE 8 : Percentage distribution of population by household expenditure class and economic status 40
TABLE 9 : Percentage distribution of population by sex and general education 41
TABLE 10 : Percentage distribution of males and females by general education 41
TABLE 11 : Percentage distribution of population by technical or professional qualification .. 42
TABLE 12 : Percentage distribution of males and females by technical or professional qualification 43
TABLE 13 : Percentage distribution of population by sex and industrial status 44
TABLE 14 : Percentage distribution of the industrial status group—children, students, old and infirm persons by age and sex 44
TABLE 15 : Percentage distribution of population by age-group and industrial status .. 45

CHAPTER FOUR

TABLE 16 : Percentage distribution of labour force by type of migrants 46
TABLE 17 : Percentage of unemployed in the labour force of different types of migrants .. 47
TABLE 18 : Percentage distribution of migrants (excluding displaced persons) in the labour force by categories of place of migration and ranges of duration of migration 48
TABLE 19 : Percentage distribution of labour force separately for males and females 49
TABLE 20 : Percentage distribution of labour force by age-group 50
TABLE 21 : Percentage distribution of population by industrial status and age-group .. 50
TABLE 22 : Percentage distribution of population by age-group separately for each industrial status 51
TABLE 23 : Percentage distribution of population by industrial status for each age-group for males and females separately 52
TABLE 24 : Percentage distribution of persons of each industrial status among the gainfully occupied in different industries 53
TABLE 25 : Percentage distribution of gainfully occupied persons by industrial status for each industry group separately 54

PAGE

TABLE 26 : Percentage distribution of gainfully occupied persons by level of monthly income .. 55

TABLE 27 : Percentage distribution of gainfully occupied persons by industry and income groups 56

TABLE 28 : Percentage distribution of employees by standard of general education and type of employer 56

TABLE 29 : Percentage distribution of employees by general education and type of employer .. 57

TABLE 30 : Percentage distribution of population by general education and industrial status .. 58

TABLE 31 : Percentage distribution of labour force between the gainfully occupied and the unemployed for each education standard 59

TABLE 32 : Percentage distribution of unemployed persons by technical or professional qualification 60

TABLE 33 : Percentage distribution of gainfully occupied persons by intensity of employment .. 61

TABLE 34 : Percentage distribution of underemployed and fully employed persons in each industry 61

TABLE 35 : Percentage distribution of underemployed persons by intensity of employment and reasons for underemployment 62

TABLE 36 : Percentage distribution of underemployed and fully employed persons by income .. 63

TABLE 37 : Income expectations of the unemployed compared with the actual incomes of the employed 64

TABLE 38 : Percentage distribution of the unemployed of each education standard by range of income expected 64

TABLE 39 : Percentage distribution of unemployed persons by types and nature of employment sought 65

TABLE 40 : Percentage distribution of unemployed persons by types and attitude to registration with employment exchanges 66

TABLE 41 : Percentage distribution of unemployed persons by general education and attitude to registration 67

TABLE 42 : Proportion of the registered to the total unemployed of each education standard .. 67

TABLE 43 : Relative importance of different reasons for non-registration (base : non-registered unemployed persons of each education standard) 68

TABLE 44 : Estimate of the educated unemployed by using employment exchange data—September 1953 70

THE NATIONAL SAMPLE SURVEY

NUMBER 8

REPORT ON PRELIMINARY SURVEY OF URBAN UNEMPLOYMENT

SEPTEMBER 1953

The Report relating to the preliminary survey of urban unemployment in September 1953 was prepared by the Indian Statistical Institute and is being published in the form in which it was submitted to the Government of India. The views contained in this report are not necessarily those of the Government of India.

CHAPTER ONE

SCOPE AND METHOD

INTRODUCTION

1.1. A countrywide survey of unemployment presents many a problem—that of concepts and definitions, of suitable sample design and of organisation, besides, of course, that of employing large resources. Before a regular estimating survey is attempted, much groundwork has to be done. The National Sample Survey accordingly decided to make a quick preliminary survey in September 1953.

1.2. The present survey was of an exploratory character. Nevertheless it has brought to light much valuable information and has for the first time furnished a fairly detailed picture of urban unemployment in the country in quantitative terms. Chapter Two gives a summary of the main results.

THE SCOPE AND DESIGN OF THE SURVEY

1.3. All towns of India with a population of 50,000 and above, except the four big cities of Calcutta, Bombay, Madras and Delhi, were brought under the geographical coverage of the survey. The four metropolitan cities of India were excluded from the enquiry as they were considered to have characteristics peculiar to themselves and were likely to be much at variance with those of the other urban areas. Another reason for their exclusion was that a special survey of unemployment in Calcutta was separately in hand and this was expected to bring out the special features of unemployment in a big city like Calcutta. The towns having a population of less than 50,000 were left out of the sample mainly on considerations of time and economy, occasioned by the fact that the NSS could spare its investigating staff for this preliminary survey only for a short period of about four weeks during the interval between the completion of its sixth round and the commencement of the seventh round. The geographical coverage of the survey nonetheless extended over 175 towns having a total population of a little over 23 million according to the 1951 Census. This represents nearly two-fifths of the total population living in towns and cities of India excluding the four big cities of Calcutta, Bombay, Madras and Delhi.

1.4. The sampling design followed the pattern of the urban sample of the NSS sixth round. The six population zones of Census 1951 were regrouped into four population areas as follows :

1. *North and North-West India*—Uttar Pradesh, Rajasthan, Punjab, PEPSU, Ajmer, Delhi, Himachal Pradesh, Bilaspur.
2. *East India*—Bihar, Orissa, West Bengal, Assam, Manipur, Tripura.
3. *Central India*—Madhya Pradesh, Madhya Bharat, Hyderabad, Bhopal, Vindhya Pradesh.
4. *South and West India*—Madras, Mysore, Travancore-Cochin, Coorg, Bombay, Saurashtra, Kutch.

1.5. The towns were classified into two size-classes, one having a population of not less than 1,00,000 and the other having a population between 50,000 and 1,00,000. The towns in each size-class were further subdivided into four groups according to the population area to which they belonged. In this way, eight strata were formed.

TABLE 1(1) : SELECTION OF SAMPLE TOWNS

population area	sample towns			
	population above 1,00,000		between 50,000-1,00,000	
	name	population (in 000)	name	population (in 000)
(1)	(2)	(3)	(4)	(5)
1. *North and North-West India* Uttar Pradesh, Rajasthan, Punjab, PEPSU, Ajmer, Delhi, Himachal Pradesh, Bilaspur	1. Rampur 2. Banaras 3. Kanpur 4. Ajmer City 5. Amritsar	1,34 3,42 6,36 1,97 3,26	15. Sambhal 16. Gonda[1]	61 33
2. *East India* Bihar, Orissa, West Bengal, Assam, Manipur, Tripura	6. Gaya 7. Tollygunge	1,34 1,50	17. Baranagar 18. Serampore	77 74
3. *Central India* Madhya Pradesh, Madhya Bharat, Hyderabad, Bhopal, Vindhya Pradesh	8. Bhopal 9. Hyderabad 10. Ujjain City	1,02 10,86 1,30	19. Kothagudem 20. Nanded	50 65
4. *South & West India* Madras, Mysore, Travancore-Cochin, Coorg, Bombay, Saurashtra, Kutch	11. Kakinada 12. Ahmedabad 13. Mangalore 14. Tiruchirapalli	1,00 7,88 1,17 2,19	21. Gadag 22. Nadiad 23. Palghat	66 63 70

---

[1] In the list obtained from the Census Superintendent, Uttar Pradesh, some time in 1951, the population of Gonda was shown as 76,331. This explains the reason for its inclusion in the list of sample towns. Later, on publication of the final figures of the 1951 Census the town was found to contain actually 32,566 people. Our estimate was, however, based on the Census Superintendent's figures.

1.6. The sampling was done in three stages. In the first stage, a few towns were selected from each stratum with probability proportional to population and with replacement. The number of sample towns selected for this survey was 23 of which 14 were drawn from towns with population above 1,00,000. For the purpose of census operation conducted in 1951, each town had been divided into a number of enumeration blocks. These blocks provided the sample units in the second stage. Each sample block was selected with probability proportional to number of households and with replacement. Ten blocks were selected from each of the sample towns having a population above 1,00,000, and eight blocks were chosen from each of the sample towns having a population between 50,000 and 1,00,000. In the third stage, a number of households were selected for investigation from each of the sample blocks. In all, 4446 sample households were investigated in the survey.

1.7. The investigation was conducted, as in other NSS enquiries, by personal visits to the households forming the ultimate sample units and the investigators collected information for all the members of the sample households by interviewing one or more members belonging to the household.

CONCEPTS AND DEFINITIONS

1.8. Before we proceed to discuss the results of the survey it will be useful to explain the more important concepts and definitions adopted in the present survey.

1.9. *Industrial status.* The term 'industrial status' is used in this survey to classify persons according to whether they are (1) gainfully occupied, (2) unemployed, seeking employment, (3) not in the labour force. Gainfully occupied persons are further classified into four sub-categories : (1) employer, (2) employee, (3) own-account worker, (4) unpaid household labour.

1.10. 'Unpaid household labour' refers to those who receive no payment in cash or kind for their services in a family trade or occupation, but who may be boarding or lodging with the household in lieu of their services. They have to be distinguished from 'family members engaged in domestic work only' who do not belong to the labour force at all and represent an important component of the category 'not in the labour force'.

1.11. An unemployed person may be either a new entrant to the labour force and seeking job for the first time or may have currently lost his employment and may be seeking one. He may, therefore, be identified either as (1) unemployed, seeking employment for the first time or as (2) unemployed, seeking employment not for the first time.

1.12. Gainfully occupied persons and unemployed persons seeking employment (either 'for the first time' or 'not for the first time') together constitute the 'labour force'.

1.13. Persons 'not in the labour force' are classified into four sub-categories : (1) family members engaged in domestic work only, (2) rentiers and pensioners, (3) beggars and persons living on charities and remittances, (4) children, students and old and infirm persons.

1.14. *The employed and the unemployed.* Any person working for gain or profit on any one day during the reference period of one month (see para 1.18) has been treated as *employed.*

1.15. A person has been considered *unemployed* if (a) he was without a job for the whole of the reference period of one month, (b) was seeking a job and (c) was in a position to accept a job if he was offered one. Further, if a person did not seek work either on account of illness or on the belief that no work was available for him or in the hope of returning to the job from which he was indefinitely laid off, he was still marked as an unemployed person. On the other hand, a person was considered employed if he had a job although he might be temporarily absent from work during the reference period on account of illness, vacation, industrial disputes or temporary lay off. Again, a person is deemed to be employed if he is only waiting to take up a new job which has been offered to him or getting ready to start a new business.

1.16. *Reference period.* The reference period for the present survey was a period of 30 days preceding the date of enquiry. During this comparatively long period of reference, it is probable that, in some cases, the 'industrial status' of a person (as defined in para 1.11) may undergo some change. For example, a person who was not in the labour force during the earlier part of the month, may enter the labour force as a gainfully occupied person in the later part of the month. In such a situation, his status was determined in accordance with a scale which gave to 'in the labour force' precedence over 'not in the labour force' and to 'gainfully occupied' precedence over 'unemployed seeking employment'. The application of this principle resulted in classifying an individual as gainfully occupied for the purpose of this survey if he was gainfully occupied for at least one day during the reference period, irrespective of whether he was gainfully occupied or unemployed during the rest of the period.

1.17. *Difficulties of comparison.* The concept of unemployment adopted in the present survey is the one advocated by the ILO. It is, however, at variance with the concept which has been adopted in several other surveys. Even the NSS special survey of unemployment in Calcutta, conducted at about the same time as the present survey, has used a different concept for determining unemployment. Any person found without employment but seeking employment on the date of the survey was deemed unemployed in that survey. The definition adopted in the present enquiry should thus give a lower estimate of unemployment than the one adopted in the Calcutta survey. To what extent the two should differ, it is not possible to say at this stage. In the NSS ninth round (August to October 1955), currently being conducted, an enquiry on unemployment all over the country has been undertaken on a large scale and one of the interesting points being investigated is to determine how far the different reference periods—a day, a week and a month—affect the estimates of unemployment.

RELIABILITY OF ESTIMATES

1.18. In a field where quantitative information is almost entirely lacking, it is not possible to check the results of the survey with existing information from other sources. The only check is afforded by a comparison of estimates obtained from independent sub-samples. In Appendix 1, a comparison has been made between estimates obtained from the two sub-samples in respect of certain important characteristics, and this indicates that the two sub-sample estimates are of the same order. In many cases, however, the estimates may not be considered satisfactory if high precision is aimed at.

Chapter Two

SUMMARY OF RESULTS

2.1. This preliminary survey of unemployment in urban areas, conducted by the National Sample Survey during September 1953, was in the nature of a try-out intended to reveal the problems and difficulties likely to arise in an investigation of this type and to enable, on the basis of experience gained, the development of suitable concepts and the framing of proper design for carrying out similar sample surveys in future with a view to providing reliable estimates. The survey covered all towns of India with a population of 50,000 and over except the four big cities of Calcutta, Bombay, Madras and Delhi. 4446 sample households drawn from 23 sample towns were visited and detailed information was collected through the usual NSS method of interview.

2.2. The survey has yielded valuable information and has, for the first time, furnished a detailed picture of urban unemployment in quantitative terms.

General characteristics of the population

2.3. The problem of employment and underemployment has been studied in this report against the background of demographic and general characteristics of the population covered in the survey. These characteristics, discussed in Chapter Three, are summarised below.

2.4. It may be explained that the estimates in this Report are being given to 3 or 4 figures as a matter only of arithmetical convention and may not justify, in many cases, the retention of more than two or even one significant figure. Paragraphs in this chapter are followed by a reference within brackets to the paragraph numbers in subsequent chapters which deal in detail with relevant topics.

2.5. *Age and sex.* 55.85 per cent (males 29.40 per cent and females 26.45 per cent) of the population belong to the age-group 16–61, which may be assumed as the working age ; 40.56 per cent (20.53 per cent males and 20.03 per cent females) belong to the age-group 0–15 and 3.59 per cent constitute the older people (62 and above). (3.4—3.5)

2.6. *Economic status.* Earners and earning dependents together constitute only one-third of the total population ; two-thirds are non-earning dependents. Among males, earners are 45.26 per cent and earning dependents 8.41 per cent : earners among females are only 5.98 per cent and earning dependents 7.54 per cent. Though persons of lower age-group constitute the bulk of non-earning dependents, the proportion of non-earning dependents in the employable age-group (16–61) is quite large (49.55 per cent) among females and not inconsiderable (17.16 per cent) even among males. (3.5 and 3.6)

2.7. *Household expenditure and economic status.* More than 60 per cent of persons belong to households having a monthly expenditure of not more than

Rs. 100 ; less than 6 per cent belong to households having monthly expenditure in excess of Rs. 300. In the very low expenditure (up to Rs. 50 per month) and high expenditure (Rs. 500 or more) households for each earner there are only two dependents on an average ; in the middle expenditure households (Rs. 100–300) there is a much lower proportion of earners, there being for each earner four dependents. (3.9—3.10)

2.8. *General education.* Illiterates constitute 57.36 per cent of the population, 'literates below matric' account for 37.86 per cent ; those who have passed the matriculation examination are only 3.17 per cent ; the intermediates are 0.75 per cent and graduates and above only 0.64 per cent. All the literates put together thus constitute only 42.42 per cent of the population, the males accounting for 27.82 per cent and the females for 14.60 per cent. Among males literates are 53.89 per cent, substantially more than those (30.18 per cent) among females. (3.15—3.16)

2.9. *Technical or professional qualification.* Persons with no technical or professional qualification constitute 84.04 per cent of the urban population. Only 14.49 per cent have some professional or technical skill without any degree or equivalent diploma and a very small number, less than two-thirds of one per cent, have higher technical or professional qualification. (3.18)

2.10. *Labour force.* 32.20 per cent of the population are found to be gainfully employed ; 2.59 per cent are unemployed ; and these together make the labour force which represents 34.79 per cent of the total population. Two-thirds of the population are thus not in the labour force. Of them, about one-third are family members engaged in domestic work and the rest are children, students, old and infirm persons and persons living on charities. (3.21)

2.11. *Industrial status and age.* One out of every twelve persons of tender age (age-group 7–15 years) is gainfully occupied. More than two-fifths of the number of persons of the age-group 16–61, which may be considered the employable age-group, are economically inactive of whom more than three-fourths are family members engaged in domestic work and the rest are mostly students but also include rentiers, pensioners, persons living on charity and old and disabled persons. Among males of this age-group nearly nine-tenths are in the labour force but among females of this age-group not even one-fifth are in the labour force and seven-tenths are engaged in domestic duties. (3.25)

THE EMPLOYED AND THE UNEMPLOYED

2.12. *Migration.* Nearly two-fifths of the labour force is composed of migrants, more than half of whom have come from rural areas and about a sixth from urban areas. A little less than one-fourth are displaced persons from Pakistan. (4.3)

2.13. The percentage of unemployed among migrants from the rural and urban areas of the same state and from rural areas of other states is of the same order and varies from 3.47 to 4.91 while for local persons it is 6.36. Migrants from urban areas of other states are unemployed to the extent of 11.43 per cent and those from East Pakistan reveal the highest percentage of 15.10. (4.5)

2.14. *Extent of unemployment.* Unemployed persons in urban areas constitute 2.59 per cent of the total population of which 1.16 per cent are seeking employment for the first time and 1.43 per cent are now out of work but had some gainful occupation before. (4.7)

2.15. Unemployed persons among males are 4.32 per cent and among females only 0.75 per cent. The unemployed males, however, represent 7.76 per cent of the male labour force and unemployed females 6.0 per cent of the female labour force. (4.8—4.9)

2.16. *Age and unemployment.* There are unemployed persons even in the age-group 7–15 years and they constitute 9.32 per cent of the labour force of this age-group. But the incidence of unemployment is highest in the age-groups 16–17 and 18–21 ; 21.64 and 20.56 per cent respectively of the labour force of these age-groups are unemployed. In the age-group 22–26 years, the proportion of the unemployed is much less (8.76 per cent) and it is still further reduced to 4.19 per cent in the age-group 27–36 years. This level is maintained till the age of 56 but there is a decline to 2.03 per cent after 57. Taking all ages together, the unemployed constitute 7.44 per cent of the labour force. (4.11)

2.17. *Pattern of employment.* Arranged in order of importance, manufacturing industries account for the largest percentage (31.87 per cent) of the gainfully occupied followed by 'other services' (22.27 per cent), distributive services (17.98 per cent), agriculture, animal husbandry, fishery and forestry (11.30 per cent), and transport, storage, communication (10.05 per cent). (4.19)

2.18. The gainfully occupied are comprised of four categories—employers, employees, own-account workers and unpaid household labour. Unpaid household labour is practically confined to the four industries—agriculture, animal husbandry, fishery and forestry (34.40 per cent) ; manufacture (30.46 per cent) ; distributive services (21.51 per cent) ; and other services (10.04 per cent). These four industries account for nearly nine-tenths of own-account workers also. Practically all the employers are also to be found therein. (4.20—4.21)

2.19. The four industries which account for nearly nine-tenths of the employees are manufacturing industries (35.05 per cent), other services (29.94 per cent) transport, storage and communication (13.06 per cent) and distributive services (8.33 per cent). (4.21)

2.20. *Employment and level of income.* Among the gainfully occupied persons 61.02 per cent have an income of less than Rs. 50 per month. 86.33 per cent earn less than Rs. 100 per month. Persons whose income exceeds Rs. 300 per month are only a little over 1 per cent. (4.25)

2.21. Incomes in some industries are better than in others. For example, in agriculture nearly 90 per cent of the gainfully occupied have incomes below Rs. 50

per month; in transport this income group accounts for slightly less than 40 per cent, and incomes are generally higher. (4.26)

2.22. *Employees and type of employer.* Employees are 54.65 per cent of the gainfully occupied persons, a little less than one-fourth (23.58 per cent) of whom are employed in the government sector. In other words, one out of every eight gainfully occupied persons is employed in the government sector. The 'government sector' includes government departments, local bodies, universities and other public institutions which derive the major part of their finances from public revenues. (4.28)

2.23. Among the employees who are illiterate, the percentage of those belonging to the government sector is 11.33, and to the non-government sector 88.67. The proportion belonging to the government sector among the 'literate but below matriculate' employees are 25.65 per cent, among matriculate and intermediate employees 55 to 56 per cent and among graduate employees 72.00 per cent, compared to the 23.58 per cent of all employees who belong to the government sector. Matriculates and above are only 6.60 per cent of the employees in the non-government sector compared to 30.56 per cent in the government sector. Illiterates account for only one-fifth among employees in the government sector. They represent nearly half of all the employees in the non-government sector. (4.28—4.30)

2.24. *Education and unemployment.* Only 8.58 per cent of the gainfully occupied persons have passed matriculation or any higher examination and illiterates (48.16 per cent) and 'literates but below matric' (43.14 per cent) are very numerous. Among the unemployed, 21.59 per cent are illiterate and 60.23 per cent are 'literate but below matric'; the educated are 17.99 per cent, showing a much higher proportion of the educated among the unemployed than among the gainfully occupied. The problem of unemployment in urban areas is seen to be much more acute for the literates than for the illiterates. Only 3.46 per cent are unemployed in the labour force belonging to the group 'illiterates'. In the group 'literate but below matric' 10.10 per cent are unemployed, and among matriculates and intermediates a little over 15 per cent are without employment. There is some improvement in respect of graduates but even among them the unemployed are no less than 9.10 per cent. (4.32—4.33)

2.25. *Technical and professional qualification and unemployment.* 61.35 per cent of the unemployed have no technical or professional qualification, 34.28 per cent have professional or technical qualification without any degree or equivalent diploma and only 4.18 per cent have been trained for some profession or technology. (4.35)

2.26. *Underemployment.* A little less than three-fourths (72.93 per cent) among the gainfully occupied persons are fully employed and the rest (26.32 per cent[1]) are underemployed to various degrees. The underemployed persons represent 8.48 per cent of the population. Among the underemployed those who have work only for 'quarter or less' of the time comprise 3.17 per cent of the population. On the assumption that these seriously underemployed persons may be considered as virtually unemployed, we get 5.8 per cent of the population either unemployed (2.59 per cent) or badly underemployed (3.17 per cent). This represents a number which is more than double of the unemployed according to the definition adopted in this Report. (4.37)

---

[1] The 'not recorded' account for the balance of 0.8 per cent. The small residual 'not recorded' percentage has been generally left out of discussion for the sake of convenience. The alternative course of distributing the 'not recorded' percentage proportionately among different categories has not been generally adopted in order to ensure comparability of figures in the text of the Report with those in the Statistical Tables.

2.27. *Reasons for underemployment.* 46.89 per cent of underemployed persons gave lack of demand as the reason for their underemployment ; 18.26 per cent were unfit to accept work because of illness ; 5.83 per cent could not work for domestic reasons and 3.40 per cent did not have opportunity to work full time on account of lack of materials and equipment. Diverse reasons accounted for the underemployment of the remaining 25.62 per cent. (4.42)

2.28. *Level of income and underemployment.* Less than 5 per cent of the underemployed have monthly incomes above Rs. 100 and only 16.89 per cent have incomes exceeding Rs. 50. As many as 60 per cent have incomes below Rs. 25. The fully employed persons are naturally better off in this respect. A little more than half of them are still having incomes of less than Rs. 50 per month but this compares favourably with more than four-fifths (83.11 per cent) of the underemployed belonging to this income group. (4.43—4.45)

2.29. *Income expectation of the unemployed.* An income up to Rs. 100 per month will meet the expectation of 85.07 per cent of the unemployed and nearly half of them will be content with an income half as much. 12.97 per cent do not expect more than Rs. 100–150 per month. Barely 2 per cent expect an earning above that range. (4.46)

2.30. Income expectations of 50 per cent of the illiterate unemployed mainly centre round the income range of Rs. 26–50 ; 22.81 per cent would place their expectation at incomes not exceeding Rs. 25 ; 23.68 per cent would expect to earn an income between Rs. 51 and Rs. 100. The only two important income levels for unemployed matriculates are Rs. 51–100 which attracts 59.41 per cent of them and Rs. 101–150, which attracts 27.54 per cent. For the graduates, the expectations are somewhat higher. Even among them an income of Rs. 51–100 may satisfy 20 per cent but half would expect an income of Rs. 101–150, and the remaining 30 per cent above Rs. 150. (4.48)

2.31. *Nature of employment sought by the unemployed.* The conclusions that may be drawn in respect of employment preference are of a broad character in view of the inadequate size of the sample which does not lend itself to meaningful analysis in respect of minute classifications. One-fourth of the unemployed prefer to work on their own account and the rest seek jobs as employees. One-third of the latter seek employment as unskilled manual workers and one-fifth as skilled manual workers. The rest, wanting non-manual employment, are equally divided between those seeking administrative and clerical work and those seeking technical and professional work. (4.49—4.50)

2.32. *Attitude to registration.* Only one-fourth of the unemployed in the urban areas have registered with the Employment Exchanges and the rest have not registered on account of various reasons, namely, ignorance, remoteness, lack of faith, or mere indifference. (4.51—4.52)

2.33. There are large differences in the matter of attitude to registration among the unemployed having different education standards. The proportion of the registered among illiterate unemployed is the lowest. Only 15 per cent of them register. This percentage goes up to 27 for the 'literate below matric', 45 for matriculates, 47 for intermediates and drops to 22 for graduates. (4.54)

2.34. Among the illiterates, failure to register is ascribable primarily to ignorance (70 per cent) ; among literates who are not matriculates, non-registration is mostly due to ignorance (53 per cent) and lack of faith (29 per cent). As educational qualification increases 'lack of faith' assumes more importance. Among graduates, for example, only 14 per cent have given ignorance about Exchange as reason for non-registration, but 57 per cent have stated that they had no faith in the efficacy of registration as a means of securing employment. (4.56)

2.35. With the help of this information, and on certain assumptions, the registration data of Employment Exchanges, which otherwise can furnish only an incomplete and distorted picture of the urban unemployment, could perhaps be made to yield useful results.

2.36. *Estimate of unemployment.* Lastly, although the present survey was not primarily designed to give reliable estimates for the urban area as a whole, and was limited in its coverage to towns of population exceeding 50,000, assuming that the broad picture is applicable also to towns of population less than 50,000 one might make a rough estimate of 2.24 million unemployed in the urban areas including the four big cities of Calcutta, Bombay, Madras and Delhi and comprising an estimated population of nearly 68 million in September 1953. (4.60—4.64)

2.37. If the severely underemployed, that is, those having employment for only 'quarter or less' of the normal time are also to be treated as virtually unemployed, nearly 2.74 million more would have to be added to the estimated number of the unemployed, giving a total estimate of the unemployed and severely underemployed as very nearly 5 million. (4.64)

2.38. The number of the unemployed having an education standard of matriculation and above, and who may be called the educated unemployed, may be estimated at 0.40 million in September 1953. (4.65)

2.39 *Provisional character of estimates.* It is necessary to bear in mind that the information presented in this Report is based ultimately on data collected from only 23 towns all over the country and this, of necessity, puts some limitation on the fully representative character of the estimates. They should, therefore, be viewed more as indicators of the order of magnitudes than as anything else having high precision. The justification for presenting them is that even such order of magnitudes is of great value in a field where there is hardly any information of a quantitative nature.

Chapter Three

GENERAL CHARACTERISTICS OF THE POPULATION

3.1. Employment and unemployment are more fruitfully studied against the background of the characteristics of the population surveyed. Information in regard to some demographic characteristics of the population, namely sex, age, economic status, industrial status, educational standard and so forth, was, therefore, collected in the survey and has been suitably tabulated.

3.2. The detailed results of the survey are given in the set of Statistical Tables which come after Chapter Four. Some of the salient features are brought out in this Chapter and the next. For facilitating reference, the tables in the text carry, along with their own numbers, a reference, within brackets, to numbers of relevant Statistical Tables.

3.3. *Age and sex.* In Table 2 below is given the percentage distribution of population by age-group and sex as obtained from the survey. The scheme of grouping of ages is different from that followed in the 1951 Census and has been adopted so as to reduce considerably the effects of the observed bias in favour of particular digits.

TABLE 2(2) : PERCENTAGE DISTRIBUTION OF POPULATION BY SEX AND AGE-GROUP

	age-group (in years)	male	female	total
	(1)	(2)	(3)	(4)
1.	0— 6	9.88	9.73	19.61
2.	7—15	10.65	10.30	20.95
3.	16—17	1.95	1.66	3.61
4.	18—21	4.23	4.23	8.46
5.	22—26	5.12	4.82	9.94
6.	27—36	7.86	6.54	14.40
7.	37—46	5.35	4.56	9.91
8.	47—56	3.60	3.27	6.87
9.	57—61	1.29	1.37	2.66
10.	62 & above	1.71	1.88	3.59
11.	total	51.64	48.36	100.00
	no. of sample persons	(10,538)	(9,867)	(20,405)

3.4. A comparison has been made in the following Table between the age distribution of population covered in the survey with the age distribution of the population in urban India according to the 1951 Census sample tabulation after adjusting for the different age-groups adopted in the present survey and the Census. The agreement is fairly good.

TABLE 3 : COMPARISON OF AGE DISTRIBUTION, SAMPLE SURVEY AND 1951 CENSUS

base : each sex total = 100

age-group		sample survey		1951 census (urban India	
		males	females	females	males
(1)		(2)	(3)	(4)	(5)
1.	0— 6	19.14	20.12	18.46	16.69
2.	7—15	20.62	21.29	21.33	19.85
3.	16—26	21.90	22.14	20.94	21.54
4.	27—36	15.21	13.53	14.76	16.38
5.	37—46	10.34	9.43	10.45	11.75
6.	47—56	6.97	6.77	7.24	7.69
7.	57 and above	5.82	6.72	6.82	6.10
8.	all ages	100.00	100.00	100.00	100.00

3.5. If we consider persons below the age of 15 years as children and assume the 'working age' as lying normally between 16 to 61 years, we may summarise the information as in Table 4 below.

TABLE 4(2) : CHILDREN, PERSONS OF WORKING AGE AND OLDER PERSONS IN THE POPULATION

age-group		percentage of total			percentage of total in each sex	
		males	females	total	females	males
(1)		(2)	(3)	(4)	(5)	(6)
1.	children : 0–15	20.53	20.03	40.56	41.41	39.76
2.	working age : 16–61	29.40	26.45	55.85	54.70	56.93
3.	older people : above 61	1.71	1.88	3.59	3.89	3.31
4.	all ages	51.64	48.36	100.00	100.00	100.00

3.6. *Economic status.* It is found that 26.26 per cent of the urban population are earners, 8.00 per cent are earning dependents, and 65.56 per cent are non-earning dependents. Among males, earners are 45.26 per cent, earning dependents are 8.41 per cent and 46.11 per cent are non-earning dependents. Among females earners are much fewer, being only 5.98 per cent; the earning dependents are 7.54 per cent and as many as 86.33 per cent are non-earning dependents. Females account for less than one-fifth of all the earners and earning dependents and among them more than half are earning dependents. The earners and earning dependents among males are 53.67 per cent and among females 13.52 per cent according to the present survey. This compares fairly well with the corresponding figures of 54.4 per cent and 11.9 per cent of 1951 Census for the urban areas of the country.

TABLE 5(3) : PERCENTAGE DISTRIBUTION OF MALES AND FEMALES BY ECONOMIC STATUS

economic status	per cent of total in each sex			per cent of total population	
	males	females	total	females	males
(1)	(2)	(3)	(4)	(5)	(6)
1. earner	45.26	5.98	26.26	2.89	23.37
2. earning dependents	8.41	7.54	8.00	3.65	4.35
3. sub-total : (1) & (2)	53.67	13.52	34.26	6.54	27.72
4. non-earning dependents	46.11	86.33	65.56	41.75	23.81
5. not recorded	0.22	0.15	0.18	0.07	0.11
6. total	100.00	100.00	100.00	48.36	51.64
no. of sample persons	(10,538)	(9,867)	(20,405)		

3.7. Non-earning dependents constitute nearly two-thirds of the population. 58.88 per cent of all non-earning dependents belong to the age-group 0–15, 37.79 per cent belong to the age-group 16–61 and only 3.33 per cent belong to the age-group 62 and above. Of the male non-earning dependents 80.23 per cent belong to the age-group 0–15, 17.16 per cent belong to the age-group 16–61 and only 2.61 per cent belong to the age-group 62 and above. Among females, only 46.71 per cent of non-earning dependents are in the age-group 0–15, about half (49.55 per cent) are in the age-group 16–61 and 3.74 per cent in the age-group 62 and above. Table 6 brings out the comparison very clearly.

TABLE 6(4–6) : PERCENTAGE DISTRIBUTION OF POPULATION, MALES AND FEMALES SEPARATELY, BY AGE-GROUP AND ECONOMIC STATUS

economic status	age-group				no. of sample persons
	0–15	16–61	62 and above	total	
(1)	(2)	(3)	(4)	(5)	(6)
			all persons		
1. earners	1.70	94.60	3.70	100.00	(5,358)
2. earning dependents	17.84	76.77	5.39	100.00	(1,632)
3. non-earning dependents	58.88	37.79	3.33	100.00	(13,377)
4. not recorded	44.75	52.62	2.63	100.00	(38)
5. total	40.56	55.85	3.59	100.00	(20,405)
			males		
6. earners	1.42	95.08	3.50	100.00	(4,769)
7. earning dependents	24.23	69.57	6.20	100.00	(887)
8. non-earning dependents	80.23	17.16	2.61	100.00	(4,859)
9. not recorded	39.11	60.89	0.00	100.00	(23)
10. total	39.76	56.93	3.31	100.00	(10,538)
			females		
11. earners	3.91	90.83	5.26	100.00	(589)
12. earning dependents	10.20	82.37	4.43	100.00	(745)
13. non-earning dependents	46.71	49.55	3.74	100.00	(8,518)
14. not recorded	53.32	40.01	6.67	100.00	(15)
15. total	41.41	54.70	3.89	100.00	(9,867)

3.8. Though persons of lower age-group constitute the bulk of non-earning dependents, the proportion of non-earning dependents in the 'working age-group' (16–61) is quite large (49.55 per cent) among females and not inconsiderable (17.16 per cent) even among males. In regard to females it may be said that most of the dependents in this age-group are engaged in domestic work. The high proportion of male dependents is accounted for almost half by students or infirm persons in this age-group and half by the unemployed. A small proportion consists also of rentiers and pensioners, persons living on charities and persons engaged in domestic work only.

3.9. *Household expenditure and economic status.* If we classify the people according to the level of monthly expenditure of the household to which they belong we get the result, summarised in Table 7 below.

TABLE 7(7b) : PERCENTAGE DISTRIBUTION OF POPULATION BY MONTHLY HOUSEHOLD EXPENDITURE CLASS

	monthly household expenditure class (Rs.)	per cent of persons	cumulative percentage[1] from	
			below	above
	(1)	(2)	(3)	(4)
1.	up to 50	24.28	100.00	24.48
2.	51—100	36.14	75.52	60.91
3.	101—200	25.17	39.09	86.29
4.	201—300	7.82	13.71	94.17
5.	301—500	4.22	5.83	98.42
6.	501-1000	1.05	1.58	99.48
7.	above 1000	0.52	0.52	100.00
8.	not recorded	0.80	—	—
9.	total	100.00		

3.10. 60.91 per cent of the persons belong to households having a monthly expenditure of less than Rs. 100, and 5.83 per cent belong to households having a monthly expenditure above Rs. 300.

3.11. An interesting feature is the much lower proportion of earners and correspondingly higher proportion of earning and non-earning dependents in the households of middle levels of expenditure compared to the households of low expenditure level on the one hand and high expenditure level on the other. This is shown in the following Table.

---

[1] After distributing proportionately in each class the 'not recorded' percentage (0.80) and cumulating from below and above.

TABLE 8(7a) : PERCENTAGE DISTRIBUTION OF POPULATION BY HOUSEHOLD EXPENDITURE CLASS AND ECONOMIC STATUS

economic status	monthly expenditure class (in Rs.)						
	up to 50	51–100	101–200	201–300	301–500	501–1000	above 1000
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. earners	33.73	25.51	22.29	20.55	21.25	31.16	32.08
2. earning dependents	9.69	8.82	6.72	5.64	5.12	1.40	10.37
3. non-earning dependents	56.16	65.60	70.83	73.81	73.63	66.04	57.55
4. not recorded	0.42	0.07	0.16	0.00	0.00	1.40	0.00
5. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(4,954)	(7,375)	(5,136)	(1,596)	(861)	(215)	(106)

3.12. In the very low expenditure (up to Rs. 50 per month) households and the high expenditure households (Rs. 500 or more per month), for every earner there are only two earning or non-earning dependents on an average, but in the middle expenditure households (Rs. 100–300 per month), for every earner, there are four earning or non-earning dependents.

3.13. *General education.* The general educational standard for the purpose of this survey has been classified into five categories: (1) illiterate, (2) literate but below matric, (3) matric, (4) intermediate, (5) graduate and above.

3.14. A person has been considered 'illiterate' if he cannot read and write any language. The category 'literate but below matric' comprises persons who satisfy this minimum condition of literacy but have not passed the matriculation or any other examination of equivalent standard. 'Matric' refers to persons who have passed matriculation or an examination of equivalent standard but have not proceeded beyond. The group 'intermediate' refers to persons who have passed the intermediate examination in arts, science or commerce. The fifth group, graduates and above, is constituted of those who hold university degrees.

3.15. The present survey has revealed that illiterates constitute 57.36 per cent of the population in towns covered by the survey, 'literate but below matric' account for 37.86 per cent, those who have only passed the matriculation examination are 3.17 per cent, the 'intermediates' are 0.75 per cent and 'graduates and above' only 0.64 per cent. For 0.22 per cent educational particulars could not be collected for one reason or other.

3.16. All the literates put together thus constitute only 42.42 per cent of the population, the males accounting for 27.82 per cent and the females for 14.60 per cent. The corresponding percentages for male and female literates according to Census 1951 Urban-Sample tabulation (Census of India paper No. 5 Table D-VIII) are 24.30 and 10.29. The Census tables do not give the information for urban areas separately in respect of detailed classification of literates by their educational standard to enable a more detailed comparison being made. The somewhat higher proportion of literates both among males and females, as found in the present survey, may be due to the fact that the Census data relate to the entire urban area which includes, besides the larger towns covered in the survey, also towns having 5000 to 50,000 people. This latter class of towns are likely to have fewer literates than the larger towns with the result that the over-all literacy in the entire area as obtained from the Census will evidently be less than that estimated from the survey.

TABLE 9(8a) : PERCENTAGE DISTRIBUTION OF POPULATION BY SEX AND GENERAL EDUCATION

general education	male	female	total	no. of sample persons
(1)	(2)	(3)	(4)	(5)
1. illiterate	23.73	33.63	57.36	(11,704)
2. literate but below matric	23.82	14.04	37.86	(7,725)
3. matric	2.73	0.44	3.17	(647)
4. intermediate	0.69	0.06	0.75	(154)
5. graduate and above	0.58	0.06	0.64	(130)
6. not recorded	0.09	0.13	0.22	(45)
7. total	51.64	48.36	100.00	(20,405)

3.17. Among males, again, 45.05 per cent are illiterate and another 46.14 per cent are literate but have not passed the matriculation or any other equivalent examination. Matriculates and intermediates account for 6.62 per cent and graduates and persons having higher general education constitute 1.13 per cent. Illiterates among females are very numerous (69.54 per cent) while literates below matric constitute 29.02 per cent. The percentage of females having a general education, matriculation or above, is only 1.16.

TABLE 10(8b) : PERCENTAGE DISTRIBUTION OF MALES AND FEMALES BY GENERAL EDUCATION

general education	male	female	total
(1)	(2)	(3)	(4)
1. illiterate	45.95	69.54	57.36
2. literate but below matric	46.14	29.02	37.86
3. matric	5.29	0.91	3.17
4. intermediate	1.33	0.14	0.75
5. graduate and above	1.13	0.11	0.64
6. not recorded	0.16	0.28	0.22
7. total	100.00	100.00	100.00
number of sample persons	(10,538)	(9,867)	(20,405)

3.18. *Technical or professional qualifications.* Technical and professional qualifications of persons have been broadly classified into four categories, namely, (a) no technical or professional qualification, (b) professional or technical skill without any degree or equivalent diploma, (c) certificate or diploma in typing, stenography, nursing, midwifery or pharmacy, (d) degree or equivalent diploma in any technology or profession.

3.19. Persons with no technical or professional qualifications constitute 84.04 per cent of the population of whom 38.54 per cent are males and 45.50 per cent are females. Only 14.49 per cent have some professional or technical skill without any degree or equivalent diploma (excluding certificate or diploma in typing, stenography, nursing, midwifery and pharmacy) and a very small number, only 0.65 per cent of the entire population, have medium or higher technical or professional qualification.

TABLE 11(8a) : PERCENTAGE DISTRIBUTION OF POPULATION BY TECHNICAL OR PROFESSIONAL QUALIFICATION

technical or professional qualification	male	female	total	number of sample persons
(1)	(2)	(3)	(4)	(5)
1. no technical or professional qualification	38.54	45.50	84.04	(17,147)
2. professional or technical skill without any degree or diploma and not covered by category 3 below	11.85	2.64	14.49	(2,957)
3. certificate or diploma in typing, stenography, nursing, midwifery or pharmacy	0.69	0.08	0.77	(159)
4. degree or equivalent diploma in any technology or profession	0.47	0.01	0.48	(97)
5. not recorded	0.09	0.13	0.22	(45)
6. total	51.64	48.36	100.00	(20,405)

3.20. The disparity between the two sexes in this respect is seen clearly in Table 12. In both sexes persons having no technical or professional qualification predominate. They account for 74.67 per cent of the males and 94.05 per cent of the females. The divergence between the sexes increases as we ascend the scale of qualifications. The higher the qualification, the greater the divergence. There are, for example, proportionately four times as many men with professional or technical skill without any degree or equivalent diploma among males as there are such women among females. The divergence is more than double for the next higher qualification and very much larger for the next following.

TABLE 12(8b) : PERCENTAGE DISTRIBUTION OF MALES AND FEMALES BY TECHNICAL OR PROFESSIONAL QUALIFICATION

technical or professional qualification	male	female	total
1. no technical or professional qualification	74.67	94.05	84.04
2. professional or technical skill without any degree or equivalent diploma and not covered by category 3 below	22.93	5.48	14.49
3. certificate or diploma in typing, stenography, nursing, midwifery or pharmacy	1.36	0.16	0.77
4. degree or equivalent diploma in any technology or profession	0.88	0.03	0.48
5. not recorded	0.16	0.28	0.22
6. total	100.00	100.00	100.00
number of sample persons	(10,538)	(9,867)	(20,405)

3.21. *Industrial status.* We may conclude this Chapter with a discussion on industrial status which describes whether a person is gainfully occupied, is unemployed and seeking employment or is not in the labour force. This will be in the nature of a brief introduction to the 'labour force' which is of main interest in the present enquiry and to a detailed consideration of which the whole of the Chaprer Four is devoted. The gainfully occupied are only 32.20 per cent of the population. The unemployed are 2.59 per cent and the two of them together constitute the 'labour force' which is 34.79 per cent of the population. The Census 1951 concept of 'working force' is somewhat different from the concept of 'labour force' adopted in the present survey. The unemployed do not figure in the 'working force' of the Census which is comprised simply of the earners and earning dependents. It does not correspond to the category 'gainfully occupied' either, inasmuch as every earner or earning dependent need not be in the labour force at all. A rentier or a beggar, for example, who maintains himself from his income is an 'earner' but he is not 'gainfully occupied'. If we add to the number of 'gainfully occupied' the number of rentiers, pensioners, and persons living on charities etc., our concept would fairly closely approximate the Census concept of 'working force'. This may be estimated from Table 13 below by adding the figures in column 4 at row 1 and row 5. As against 33.6 per cent thus arrived at, we have the Census figure of 'working force' as 34.7 per cent of the total population for the entire urban area which includes towns of population size 5000–50,000 also.

3.22. Persons 'not in labour force' constitute 64.66 per cent of the population out of which family members engaged in domestic work account for 22.28 per cent; children and students, old and infirm persons 40.95 per cent; rentiers and pensioners 0.61 per cent, beggars 0.82 per cent.

TABLE 13(9) : PERCENTAGE DISTRIBUTION OF POPULATION BY SEX AND INDUSTRIAL STATUS

industrial status	males	females	total
(1)	(2)	(3)	(4)
Persons in labour force			
1. gainfully occupied	51.33	11.74	32.20
2. unemployed	4.32	0.75	2.59
3. sub-total	55.65	12.49	34.79
Persons not in labour force			
4. family members engaged in domestic work only	0.62	45.42	22.28
5. rentiers, pensioners, beggars etc.	1.32	1.55	1.43
6. children, students, old and infirm persons	41.86	39.98	40.95
7. sub-total	43.80	86.95	64.66
8. not recorded	0.55	0.56	0.55
9. total	100.00	100.00	100.00
number of sample persons	(10,538)	(9,867)	(20,405)

3.23. Persons 'not in the labour force' are proportionately twice as many among females (86.95 per cent) as among males (43.80 per cent). This is mainly due to the very high proportion of females (45.42 per cent) who are engaged in domestic work only ; the corresponding proportion for males is only 0.62 per cent.

3.24. In the rather mixed category of 'children and students, old and infirm persons' about 92 per cent are children and students of which 47 per cent belong to the age-group 0–6 years, 40 per cent are boys and girls of age 7–15 years and 5 per cent of age 16–21 years. About 6 per cent are either 57 years old or have crossed that limit.

TABLE 14(10–12) : PERCENTAGE DISTRIBUTION OF THE INDUSTRIAL STATUS GROUP 'CHILDREN, STUDENTS, OLD AND INFIRM PERSONS' BY AGE AND SEX

sex	age-group					
	0–6	7–15	16–21	22–56	57 & above	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. males	44.96	41.44	7.87	1.86	3.87	100.00
2. females	49.71	36.60	2.46	3.25	7.98	100.00
3. total	47.20	39.16	5.31	2.52	5.81	100.00

3.25. The summary results which are given below in Table 15 reveal some striking features. On the one hand, as many as one out of every 12 persons of tender age (age-group 7–15) are gainfully employed, economic circumstances making it perhaps obligatory on many children to take up work for a living ; on the other hand, quite a large number of persons of the age-group 16–61, that is, the working age-group, are economically inactive, a large proportion of them being females who are kept busy in domestic activities.

TABLE 15(10a) : PERCENTAGE DISTRIBUTION OF POPULATION BY AGE-GROUP AND INDUSTRIAL STATUS

industrial status	age-group										
	0-6	7-15	16-17	18-21	22-26	27-36	37-46	47-56	57-61	62 & above	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1. gainfully occupied persons	0.07	8.66	27.08	36.71	52.39	61.50	61.94	56.46	44.39	28.37	32.20
2. unemployed persons seeking employment	0.00	0.89	7.48	9.50	5.03	2.69	2.28	2.49	0.92	0.55	2.59
3. family members engaged in domestic work	0.35	13.10	33.47	40.59	39.58	33.19	31.96	27.80	21.36	13.10	22.28
4. other persons not in the labour force	98.55	76.65	31.70	12.85	2.66	2.21	3.47	13.04	32.96	57.57	42.38
5. not recorded	1.03	0.70	0.27	0.35	0.34	0.41	0.35	0.21	0.37	0.41	0.55
6. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
number of sample persons	(4,002)	(4,274)	(735)	(1,727)	(2,029)	(2,938)	(2,021)	(1,403)	(543)	(733)	(20,405)

3.26. The preceding paragraphs give a brief account of some of the characteristics of the general population. The Chapter which follows deals more specifically with employment and unemployment in their diverse aspects.

Chapter Four

THE EMPLOYED AND THE UNEMPLOYED

4.1. In this Chapter we propose to present in some detail the information on employment and unemployment. The topics covered are trends of migration and their influence on employment, the structure of the labour force, the pattern of employment and income in different industries, the extent of unemployment and underemployment, the qualifications, expectations and preferences of the unemployed, the attitude of the unemployed to registration with Employment Exchanges. The Chapter concludes with an estimate of the number unemployed in the urban areas.

TABLE 16(15a) : PERCENTAGE DISTRIBUTION OF LABOUR FORCE BY TYPE OF MIGRANTS

type of migrants or non-migrants	gainfully occupied	un-employed	total
(1)	(2)	(3)	(4)
non-migrants			
1. local persons	54.21	45.84	53.59
2. temporary visitors	4.08	4.73	4.13
sub-total	58.29	50.57	57.72
migrants from			
3. rural area of the same state	18.27	11.74	17.78
4. urban area of the same state	6.18	3.60	5.99
5. rural area of different states	5.08	2.27	4.88
6. urban area of different states	0.94	1.52	0.99
7. East Pakistan as displaced persons	4.96	10.98	5.41
8. West Pakistan as displaced persons	3.84	5.30	3.94
9. other places	0.90	0.57	0.87
sub-total	40.17	35.98	39.86
10. not recorded	1.54	13.45	2.42
11. total	100.00	100.00	100.00
no. of sample persons	(6,569)	(528)	(7,097)

4.2. *Migration.* A large proportion of persons constituting the labour force consists of those who do not belong originally to the town of their present residence. They have come to the town at various periods with the intention of staying there for more than a year. They are termed migrants as distinct from temporary visitors who may have come to a town purely for a short visit and with no intention of staying for at least a year. These migrants have come, over many years, from various parts such as rural area of the same state, urban area of the same state, rural or urban area of a different state or from outside the country. A special category of migrants is composed of those persons who have come as displaced persons from East or West Pakistan. The following Table shows the proportion of the local persons, the temporary visitors and the migrants of different categories in the labour force of our urban areas.

4.3. Nearly 40 per cent of the labour force is composed of migrants, more than half of whom have come from rural areas, about a sixth from urban areas and a little less than one-fourth are displaced persons from Pakistan.

4.4. This description is more or less applicable to the gainfully occupied group but has to be substantially modified in the case of the unemployed. Among the unemployed also the migrants are almost two-fifths, of which two-fifths are from rural areas and between half and two-fifths are displaced persons.

TABLE 17(15b) : PERCENTAGE OF UNEMPLOYED IN THE LABOUR FORCE OF DIFFERENT TYPES OF MIGRANTS

type of migrants or non-migrants	percentage of unemployed in the labour force of each migrant or non-migrant class		
	for the first time	not for the first time	total
(1)	(2)	(3)	(4)
non-migrants			
1. local persons	2.71	3.65	6.36
2. temporary visitors	4.09	4.44	8.53
sub-total	2.81	3.71	6.52
migrants			
3. rural area of the same state	2.06	2.85	4.91
4. urban area of the same state	1.65	2.82	4.47
5. rural area of different states	1.45	2.02	3.47
6. urban area of different states	7.14	4.29	11.43
7. East Pakistan as displaced persons	7.81	7.29	15.10
8. West Pakistan as displaced persons	3.57	6.43	10.00
9. other places	1.61	3.23	4.84
sub-total	2.97	3.75	6.72
10. not recorded	22.09	19.19	41.28
11. total	3.34	4.10	7.44
number of sample persons	(237)	(291)	(528)

4.5. The question arises as to how these various categories of migrants have fared in regard to employment. It is interesting to find that the percentage of unemployed among the migrants from the rural and urban areas of the same state and from rural areas of other states varies from 3.47 to 4.91 and is less than the percentage of unemployed among the local persons, which is 6.36 per cent, and about half for the entire labour force. Compared to the average rate of 7.44 per cent for the entire labour force, migrants of this class are much better off. On the other hand, migrants from other urban areas are much worse off with unemployment of 11.43 per cent. Worst are displaced persons from East Pakistan showing an unemployment percentage of 15.10. Displaced persons from West Pakistan fare better, but with unemployment at 10.00 per cent, the situation is far from satisfactory. Table 17 (see page 23) presents this information for the different types of migrants or non-migrants and separately for the unemployed persons, seeking employment 'for the first time' and 'not for the first time'.

4.6. To study the normal trends in migration, one should take into account the situation created by the influx of displaced persons into India from East and West Pakistan in the years following the Partition. In the following Table is shown a percentage distribution of migrants (excluding displaced persons) in the labour force of urban areas by place of migration and by range of duration of migration.

TABLE 18(18a) : PERCENTAGE DISTRIBUTION OF MIGRANTS (EXCLUDING DISPLACED PERSONS) IN THE LABOUR FORCE BY CATEGORIES OF PLACE OF MIGRATION AND RANGES OF DURATION OF MIGRATION

category of place of migration	ranges of duration of migration				
	1 year-2 years	2 years-4 years	4 years-6 years	6 years-10 years	10 years-15 years
(1)	(2)	(3)	(4)	(5)	(6)
1. rural area of the same state	51.50	60.49	52.38	64.33	59.91
2. urban area of the same state	23.95	18.49	24.66	17.97	19.40
3. rural area of different states	20.36	18.00	15.14	14.90	11.64
4. urban area of different states	4.19	2.50	4.33	2.24	3.45
5. other places	0.00	0.52	3.49	0.56	5.60
6. total	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(167)	(200)	(231)	(356)	(232)

4.7. *Unemployment and sex.* Unemployed persons seeking employment in the urban areas (under survey) constitute 2.59 per cent of the total population of which 1.16 per cent are fresh additions to the unemployed and 1.43 per cent are out of work now but had some gainful occupation before.

4.8. The proportion of unemployed persons among males (4.32 per cent) is much higher than among females (0.75 per cent). It should not, however, be inferred that employment opportunities are better for females than these are for males. The reason for the lower percentage of unemployment among females may be found in the fact that a much smaller proportion of females is in the female labour force.

4.9. The disparity between the two sexes in regard to unemployment is seen to be much less if the proportion of the labour force in each sex, unemployed at present, is taken as the index of unemployment. The unemployed males represent 7.76 per cent of the male labour force; and the unemployed females account for 6.00 per cent of the female labour force. Taking both sexes together 7.44 per cent of the labour force is unemployed.

TABLE 19(9) : PERCENTAGE DISTRIBUTION OF LABOUR FORCE SEPARATELY FOR MALES AND FEMALES

labour force	male	female	total
(1)	(2)	(3)	(4)
gainfully employed persons			
1. employer	1.49	0.56	1.35
2. employee	50.64	50.28	50.56
3. own-account worker	34.36	25.22	32.77
4. unpaid household labour	5.75	17.94	7.88
sub-total	92.24	94.00	92.56
unemployed persons seeking employment			
5. for the first time	3.47	2.80	3.33
6. not for the first time	4.29	3.20	4.11
sub-total	7.76	6.00	7.44
7. total	100.00	100.00	100.00
number of sample persons	(5,865)	(1,232)	(7,097)

4.10. A point of some interest is the much larger proportion of the unpaid household labour among females (17.94 per cent) than among males (5.75 per cent). This category is three times as numerous among females than among males.

4.11. *Age and unemployment.* The impact of unemployment is more marked on persons of lower age-groups. This is seen clearly in Table 20 which gives, for each age-group, the proportion of the gainfully occupied and the unemployed in the labour force. It will be readily seen that incidence of unemployment is highest in the age-groups 16–17 and 18–21; more than 1 out of every 5 (20.56–21.64 per cent) in the labour force in these age-groups is in search of a job. In the age-group 22–26 years the proportion of the unemployed in much less (8.76 per cent) and it is still further sharply reduced to less than half as much (4.19 per cent) in the next age-group of 27–36. Till the age of 56 this level is more or less maintained but there is again a decline to 2.03 per cent for the age-group 57–61 and 1.90 per cent for the age-group 62 and above.

TABLE 20(10) : PERCENTAGE DISTRIBUTION OF LABOUR FORCE BY AGE-GROUP

industrial status	age-group										
	0-6	7-15	16-17	18-21	22-26	27-36	37-46	47-56	57-61	62 & above	total
1. gainfully occupied persons	100.00	90.68	78.36	79.44	91.24	95.81	96.45	95.78	97.97	98.10	92.56
2. unemployed persons seeking employment	0.00	9.32	21.64	20.56	8.76	4.19	3.55	4.22	2.03	1.90	7.44
3. total labour force	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

4.12. It is not without interest that even boys and girls of the age-group 7–15 have entered the labour force and they account for 12.24 per cent of the unemployed persons seeking employment for the first time. Nearly three-fourths (72.56 per cent) of all the unemployed who seek employment for the first time are below the age of 21 and barely 10 per cent are more than 26 years of age. This may be contrasted with the unemployed who seek employment not for the first time. Among them persons below the age of 21 are only 29.21 per cent and 49.83 per cent are older than 26 years.

TABLE 21(10b) : PERCENTAGE DISTRIBUTION OF POPULATION BY INDUSTRIAL STATUS AND AGE-GROUP

	age-group	industrial status					total
		gainfully occupied	unemployed seeking employment		not in the labour force	not recorded	
			for the first time	not for the first time			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	0— 6	0.05	0.00	0.00	30.00	36.28	19.61
2.	7—15	5.63	12.24	3.09	29.07	26.55	20.95
3.	16—17	3.03	17.72	4.47	3.63	1.77	3.61
4.	18—21	9.65	42.60	21.65	6.99	5.31	8.46
5.	22—26	16.18	17.30	20.96	6.49	6.19	9.94
6.	27—36	27.51	5.49	22.68	7.88	10.62	14.40
7.	37—46	19.06	3.38	13.06	5.43	6.20	9.91
8.	47—56	12.06	0.85	11.34	4.34	2.66	6.87
9.	57—61	3.67	0.42	1.37	2.24	1.77	2.66
10.	62 & above	3.16	0.00	1.38	3.93	2.65	3.59
11.	total	100.00	100.00	100.00	100.00	100 00	100.00
no. of sample persons		(6,569)	(237)	(291)	(13,195)	(113)	(20,405)

4.13. The features noted above, though generally true of the entire population, are more pronounced among males than among females. The main difference in the pattern is that among females a large majority are family members engaged in domestic work.

4.14. *'Potential labour force'*. One might consider all persons belonging to the age-group 16–61, as comprising what may be called 'potential labour force'.

4.15. From the summary Table given below, it is seen that this 'potential labour force' accounts for 55.85 per cent of the population. 91.16 per cent of gainfully employed persons and 92.04 per cent of unemployed persons belong to this age-group while 59.07 per cent of persons not in the labour force are seen concentrated in the age-group 0–15.

TABLE 22(10) : PERCENTAGE DISTRIBUTION OF POPULATION BY AGE-GROUP SEPARATELY FOR EACH INDUSTRIAL STATUS

industrial status	age-group			total	no. of sample persons
	0–15	16–61	62 and above		
(1)	(2)	(3)	(4)	(5)	(6)
1. gainfully occupied persons	5.68	91.16	3.16	100.00	(6,569)
2. unemployed persons seeking employment	7.20	92.04	0.76	100.00	(528)
3. persons not in the labour force	59.07	37.00	3.93	100.00	(13,195)
4. not recorded	62.83	34.52	2.65	100.00	(113)
5. total	40.56	55.85	3.59	100.00	(20,405)

4.16. Table 23 shows that 94.16 per cent of persons of the age-group 0–15 do not belong to the labour force, but 4.98 per cent have already entered the labour force.

4.17. Of persons belonging to the age-group 16–61, the age-group of potential labour force, 56.81 per cent are in the labour force and of the 42.85 per cent who are not in the labour force only one in eight is a male. The rest are females, nearly nine-tenths of whom are engaged in domestic work only.

TABLE 23(11, 12) : PERCENTAGE DISTRIBUTION OF POPULATION BY INDUSTRIAL STATUS FOR EACH AGE-GROUP FOR MALES AND FEMALES SEPARATELY

industrial status	males				females				all persons			
	age-group				age-group				age-group			
	0–15	16-61	62 & above	total	0–15	16–62	62 & above	total	0–15	16–61	62 & above	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	9)	(10)	(11)	(12)	(13)
1. gainfully occupied	6.54	82.78	48.71	51.33	2.42	18.94	9.90	11.74	4.51	52.55	28.37	32.20
2. unemployed seeking employment	0.81	6.97	0.86	4.32	0.10	1.26	0.26	0.75	0.47	4.26	0.55	2.59
3. sub-sotal : persons in labour force	7.35	89.75	49.57	55.65	2.52	20.20	10.16	12.49	4.98	56.81	28.92	34.79
4. family persons engaged in domestic work	0.72	0.53	0.86	0.62	13.31	71.24	24.22	45.42	6.93	34.02	13.10	22.28
5. rentiers and pensioners	0.00	1.00	9.17	0.87	0.00	0.56	0.78	0.33	0.00	0.79	4.77	0.61
6. persons living on charities and remittances	0.10	0.42	5.16	0.45	0.00	1.83	5.47	1.22	0.04	1.09	5.32	0.82
7. old and infirm persons, children and students	90.95	7.99	34.67	41.86	83.33	5.80	59.11	39.98	87.19	6.95	47.48	40.95
8. sub-total : not in labour force	91.77	9.94	49.86	43.80	96.64	79.43	89.58	86.95	94.16	42.85	70.67	64.66
9. not recorded	0.88	0.31	0.57	0.55	0.84	0.37	0.26	0.56	0.86	0.34	0.41	0.55
10. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(4,190)	(5,999)	(349)	(10,538)	(4,086)	(5,397)	(384)	(9,867)	(8,276)	(11,396)	(733)	(20,405)

Among males of the age-group 16-61, 82.78 per cent are gainfully occupied, 6.97 per cent are unemployed seeking work and less than 10 per cent are not in the labour force, a majority of whom being students. Among females of this age-group, the gainfully occupied are only 18.94 per cent, the unemployed 1.26 per cent and nearly 80 per cent are not in the labour force.

4.18. *Pattern of employment.* Table 24 shows how the different components of the gainfully occupied group, namely employers, employees, own-account workers, and unpaid household labour, are distributed in different industries.

TABLE 24(20b) : PERCENTAGE DISTRIBUTION OF PERSONS OF EACH INDUSTRIAL STATUS AMONG THE GAINFULLY OCCUPIED IN DIFFERENT INDUSTRIES

industry	industrial status				
	employer	employee	own-account worker	unpaid household labour	all gainfully occupied
(1)	(2)	(3)	(4)	(5)	(6)
1. agriculture, animal husbandry, forestry, fishing, hunting	23.16	5.07	14.83	34.40	11.30
2. mining and quarrying	0.00	3.20	0.17	0.00	1.81
3. manufactures	38.95	35.05	27.04	30.46	31.87
4. construction	1.05	3.18	3.40	0.54	3.00
5. electricity, water, gas and sanitary services	0.00	1.98	0.56	0.72	1.34
6. distributive services	28.42	8.33	31.60	21.51	17.98
7. transport, storage and communication	0.00	13.06	8.08	0.54	10.05
8. other services not covered above	7.37	29.94	13.97	10.04	22.27
9. not recorded	1.05	0.19	0.30	1.79	0.38
10. total	100.00	100.00	100.00	100.00	100.00

4.19. According to their importance for providing employment to the urban population, the industries can be arranged as follows, starting with the one contributing to the maximum employment :

1. manufactures	... 31.87 per cent
2. services not specifically covered	... 22.27 per cent
3. distributive services	... 17.98 per cent
4. agriculture, animal husbandry, fishery and forestry	... 11.30 per cent
5. transport, storage and communication	... 10.05 per cent
6. all others	... 6.53 per cent

4.20. Unpaid household labour is practically confined to four industries: agriculture, animal husbandry, forestry and fishing (34.40 per cent), manufactures (30.46 per cent), distributive services (21.51 per cent), and other services (10.04 per cent). These four industries account for nearly nine-tenths of the own-account workers also.

4.21. The four industries which account for nearly nine-tenths of the employees are manufactures (35.05 per cent), other services (29.94 per cent), transport, storage and communication (13.06 per cent) and distributive services (8.33 per cent). Practically all the employers are to be found in the four industries, namely, manufactures (38.95 per cent), distributive services (28.42 per cent), agriculture etc. (23.16 per cent) and other services (7.37 per cent).

4.22. A further analysis of the distribution of the gainfully occupied persons shows that the proportion of employers is very small among them while that of employees and own-account workers is large. The following Table shows for different industries separately the industrial status of employed persons.

TABLE 25(20a) : PERCENTAGE DISTRIBUTION OF GAINFULLY OCCUPIED PERSONS BY INDUSTRIAL STATUS FOR EACH INDUSTRY GROUP SEPARATELY

industry	employers	employees	own-account workers	unpaid house-hold labour	total	no. of sample persons
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. agriculture, animal husbandry, forestry, fishing and hunting	2.96	24.53	46.63	25.88	100.00	(742)
2. mining and quarrying	0.00	96.64	3.36	0.00	100.00	(119)
3. manufacture	1.77	60.07	30.04	8.12	100.00	(2,094)
4. construction	0.51	57.87	40.10	1.52	100.00	(197)
5. electricity, water, gas and sanitary services	0.00	80.68	14.77	4.55	100.00	(88)
6. distributive services	2.29	25.32	62.23	10.16	100.00	(1,181)
7. transport, storage and communication	0.00	71.07	28.48	0.45	100.00	(660)
8. other services not covered above	0.48	73.48	22.21	3.83	100.00	(1,463)
9. not recorded	4.00	28.00	28.00	40.00	100.00	(25)
10. total	1.45	54.65	35.41	8.49	100.00	(6,569)

4.23. It will be seen that in distributive services (62.23 per cent) and in agriculture etc. (46.63 per cent), own-account workers are most numerous. In all other industries, however, the proportion of employees to other types of workers is high.

4.24. 'Mining and quarrying', 'electricity, water, gas and sanitary services', 'transport, storage and communication' show no employers at all. This is explained by the fact that the 'employers' of these industries are companies, corporations etc., which are of a non-household character.

4.25. *Distribution of gainfully occupied persons by industry and income group.* If we study the distribution of gainfully occupied persons by industry and income groups we find that more than 61 per cent of employed persons have an income of less than Rs. 50 per month. Persons having an income of less than Rs. 100 per month are 86.33 per cent of all the employed. Only 5.71 per cent earn in excess of Rs. 150 per month. Incomes exceeding Rs. 300 per month are only a little over 1 per cent. The cumulative percentages given in the last two columns of Table 26 bring out this point very clearly.

TABLE 26(21b) : PERCENTAGE DISTRIBUTION OF GAINFULLY OCCUPIED PERSONS BY LEVEL OF MONTHLY INCOME

income group (Rs.)	per cent of gainfully occupied	cumulative percentage[1]	
		from below	from above
(1)	(2)	(3)	(4)
1. upto 25	32.94	100.00	35.10
2. 26— 50	24.33	64.90	61.02
3. 51— 100	23.75	38.98	86.33
4. 101— 150	7.47	13.67	94.29
5. 151— 200	2.16	5.71	96.59
6. 201— 300	1.96	3.41	98.68
7. 301— 500	0.68	1.32	99.40
8. 501—1000	0.26	0.60	99.67
9. above 1000	0.30	0.33	100.00
10. not recorded	6.15		
11. total	100.00		

4.26. Incomes are better in some industries than in others. For example, in agriculture, 88.54 per cent of the employed have incomes below Rs. 50 per month. In transport this income group accounts for only 39.82 per cent ; this is proportionately less than half as many as in agriculture.

4.27. The following Table gives the income pattern of employees from each industry separately.

---

[1] After distributing the 'not recorded' percentage (6.15) proportionately over different income groups, and cumulating from below and from above.

TABLE 27(21b) : PERCENTAGE DISTRIBUTION OF GAINFULLY OCCUPIED PERSONS BY INDUSTRY AND INCOME GROUPS

industry	income group (Rs. per month)[1]						total
	up to 25	26–50	51–100	101–150	151–300	above 300	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. agriculture, animal husbandry, forests, fishing	72.31	16.23	7.22	1.31	2.29	0.64	100.00
2. mining and quarrying	9.32	74.58	12.71	3.39	0.00	0.00	100.00
3. manufactures	34.92	23.59	30.97	7.03	2.66	0.80	100.00
4. construction	47.63	29.32	15.71	4.19	2.10	1.05	100.00
5. electricity, water, gas etc.	28.57	38.09	19.05	10.72	3.57	0.00	100.00
6. distributive services	27.66	29.74	24.09	8.66	6.29	3.56	100.00
7. transport, storage etc.	18.68	21.14	37.67	16.08	5.97	0.46	100.00
8. other services not recorded above	33.49	27.24	23.11	8.60	6.34	1.22	100.00

4.28. *Distribution of employees by type of employer.* We have seen that 54.65 per cent of the gainfully occupied persons are employees. A little less than one-fourth of them are employees in the government sector. This means that roughly one out of every eight persons is an employee in the government sector in the urban areas covered by the survey. Table 28 points to the higher level of general education of employees in the government sector.

TABLE 28(22) : PERCENTAGE DISTRIBUTION OF EMPLOYEES BY STANDARD OF GENERAL EDUCATION AND TYPE OF EMPLOYER

general education	percentage in each education group[2]			no. of sample persons
	government sector	non-government sector	total	
(1)	(2)	(3)	(4)	(5)
1. illiterate	11.33	88.67	100.00	(1,550)
2. literate but below matric	25.65	74.35	100.00	(1,606)
3. matric	55.97	44.03	100.00	(295)
4. intermediate	56.45	43.55	100.00	(62)
5. graduate	72.00	28.00	100.00	(75)
6. total	23.58	76.42	100.00	(3,588)

---

[1] After distributing the 'not recorded' percentage of Table 21(b) of Statistical Tables proportionately over different income groups.

[2] After distributing proportionately between 'government' and 'non-government' the 'not recorded' percentage in each row of Table 22(b) of Statistical Tables.

4.29. If the figures in each row of column 2 are divided by the figure at row 6 and the ratio is expressed as a percentage, we will get the per cent relatives of column 5 in Table 28 above. The significance of these per cent relatives may be easily understood. If the nature of employment in the government sector and non-government sector were such that the over-all ratio of the number of employees in the two sectors was maintained for employees of the different general education standard, all figures in column 5 would be simply 100. To the extent, the relative proportion for employees of a particular education standard is higher or lower the corresponding per cent relative will be more or less than 100. It is clear from the table that as the standard of education becomes higher, relatively much larger proportion belongs to government sector. We find that illiterate employees in government sector are less than half (48 per cent) of what their proportion might be if no selective principle came into operation. The 'literate but below matric' have received just about their due according to this criterion. The matric and intermediates are nearly two and half times over-represented (237 per cent) and the relative proportion of graduates is even higher (305 per cent). Even with such over-representation of educated persons as employees in government sector, 69.44 per cent of employees in government sector are still illiterate (20.53 per cent) and 'literate below matric' (48.91 per cent), 24.04 per cent are matriculates and intermediates and only 6.52 per cent are graduates.

4.30. In the non-government sector the proportion of the educated is very much smaller. Matriculates and above are only 6.60 per cent (as against 30.56 per cent in the government sector). 95.33 per cent are either illiterate or 'literate below matric'.

TABLE 29(22a) : PERCENTAGE DISTRIBUTION OF EMPLOYEES BY GENERAL EDUCATION AND TYPE OF EMPLOYER

general education	types of employer			
	government	non-government	not recorded	total
(1)	(2)	(3)	(4)	(5)
1. illiterate	20.53	49.57	63.29	43.17
2. literate but below matric	48.91	43.76	34.18	44.73
3. matric	19.81	4.81	2.53	8.22
4. intermediate	4.23	1.01	0.00	1.73
5. graduate and above	6.52	0.78	0.00	2.09
6. not recorded	0.00	0.07	0.00	0.06
7. total	100.00	100.00	100.00	100 00
8. number in the sample	(828)	(2,683)	(79)	(3,590)

4.31. It is likely that this higher educational level of employees in the government sector is due more to the difference in the type of industry which figures in the government sector and requires a higher standard of education than to any difference between the government and non-government sector as such. It would be interesting to study the point in greater detail on the basis of an adequate sample. The fact remains, however, that employees in the government sector do have a larger proportion of educated among them than the employees in the non-government sector.

4.32. *Education standard and unemployment.* We have already discussed in Chapter Three the education standard of the general population. It was seen that illiterates and 'literates below matric' constitute more than 95 per cent of the population and not even 5 per cent have passed matriculation or any higher examination entitling them to be called educated. The proportion of educated is somewhat higher (8.58 per cent) among the gainfully occupied, but illiterates (48.16 per cent) and literate non-matriculates (43.14 per cent) still remain numerous. The pattern is markedly different in the case of the unemployed, among whom the illiterates are only 21.59 per cent, the barely literates constitute the bulk (60.23 per cent) and the proportion of the educated is more than double (17.99 per cent) the percentage among the gainfully occupied. The comparison is brought out in the following Table.

TABLE 30(13b) : PERCENTAGE DISTRIBUTION OF POPULATION BY GENERAL EDUCATION AND INDUSTRIAL STATUS

general education	industrial status				total
	gainfully employed persons	unemployed persons	not in the labour force	not recorded	
(1)	(2)	(3)	(4)	(5)	(6)
1. illiterate	48.16	21.59	63.26	69.93	57.36
2. literate but below matric	43.14	60.23		25.66	37.86
3. matric	5.69	13.07	1.52	2.65	3.17
4. intermediate	1.37	3.03	0.36	0.88	0.75
5. graduate and above	1.52	1.89	0.15	0.00	0.64
6. not recorded	0.12	0.19	0.27	0.88	0.22
7. total	100.00	100.00	100.00	100.00	100.00
8. no. of sample persons	(6,569)	(528)	(13,195)	(113)	(20,405)

4.33. We may now concentrate attention on the characteristics of our labour force in this respect.

TABLE 31(13) : PERCENTAGE DISTRIBUTION OF LABOUR FORCE BETWEEN THE GAINFULLY OCCUPIED AND THE UNEMPLOYED FOR EACH EDUCATION STANDARD

general education	labour force		total
	gainfully occupied	unemployed persons	
(1)	(2)	(3)	(4)
1. illiterate	96.54	3.46	100.00
2. literate but below matric	89.90	10.10	100.00
3. matric	84.43	15.57	100.00
4. intermediate	84.92	15.08	100.00
5. graduate	90.90	9.10	100.00
6. total	92.56	7.44	100.00

4.34. This Table clearly brings to light an important feature, namely, the problem of unemployment in urban areas is much more acute for literates than for illiterates ; and among literates, the incidence of unemployment is highest for the matriculates and intermediates, being one and half times as much for the 'literates below matric' on the one hand as the graduates on the other. It is seen that only 3.46 per cent are unemployed in the labour force belonging to the group 'illiterates'. In the next higher group, 'literate but below matric', the unemployed are proportionately three times more numerous (10.10 per cent). The situation is much worse for the 'matric' and 'intermediate' groups wherein approximately 15 to 16 per cent of the labour force are without employment. This implies an incidence of unemployment which is more than double the average incidence for the entire labour force comprising all categories of education. The graduates are comparatively well off, only 9.10 per cent of their labour force being without employment.

4.35. *Technical and professional qualifications of the unemployed.* The following Table shows the distribution of unemployed persons by technical or professional qualifications. The unemployed have been distinguished in two categories namely unemployed seeking employment 'for the first time' and unemployed seeking employment 'not for the first time'. In the first category, 73.43 per cent have no technical or professional qualification, 24.05 per cent have some professional or technical skill without any degree or equivalent diploma and 2.52 per cent have been trained for some profession. The second category is divided almost equally between those having no technical or professional qualification and those having some. On the whole, two-fifths of the unemployed do have some professional and technical skill, but persons having training in any profession or technology are very few indeed and, to be precise, they constitute barely 4 per cent of the total.

TABLE 32(26) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY TECHNICAL OR PROFESSIONAL QUALIFICATION

technical or professional qualification	unemployed persons seeking employment		
	for the first time	not for the first time	total
(1)	(2)	(3)	(4)
1. no technical or professional qualification	73.43	51.56	61.35
2. professional or technical skill without any degree or equivalent diploma and not covered by 4	24.05	42.61	34.28
3. certificate or diploma in typing, stenography, nursing, midwifery or pharmacy	1.68	3.78	2.85
4. degree or equivalent diploma in any technology or profession	0.84	1.71	1.33
5. not recorded	0.00	0.34	0.19
6. total	100.00	100.00	100.00
no. of sample persons	(237)	(291)	(528)

4.36. *Underemployment.* An attempt was made in the survey to appraise roughly the extent of underemployment. It is difficult to devise an operational concept which will enable a quantitative evaluation of the intensity of employment. To be satisfactory, it must take into account a number of complex factors, such as the normal working days per year, the normal hours of work per day, the average income etc., in a particular occupation. The observed nature and volume of work actually handled by a person has to be related to this information to provide an idea of the intensity of employment. It has not been found easy to secure suitable data for this purpose. In the present enquiry, however, the procedure adopted was quite simple, though crude. The investigators were instructed to explain to the household member concerned that he should himself assess, approximately, his own intensity of employment during the reference period, keeping in mind his actual working days and working hours per day and relating them to the number of working days and working hours normal in his type of occupation. The intensity was simply recorded as (a) much less than quarter, (b) quarter, (c) half, (d) three quarters and (e) full, and that category was adopted which seemed to describe his evaluation to the best approximation. It may be pointed out that the basis of evaluation is the subjective judgment of the person concerned. This assessment must, therefore, be accepted with caution but in the absence of other more reliable data it is not without value.

4.37. The broad conclusion is that a little less than three-fourths of the gainfully occupied persons are fully employed, and the remaining are underemployed to various degrees.

TABLE : 33(23) : PERCENTAGE DISTRIBUTION OF GAINFULLY OCCUPIED PERSONS BY INTENSITY OF EMPLOYMENT

intensity of employment	per cent of gainfully occupied
(1)	(2)
1. quarter or less	9.83
2. half	8.76
3. three quarters	7.73
4. underemployed	26.32
5. fully employed	72.93
6. not recorded	0.75
7. total	100.00
8. no. of sample persons	(6,569)

4.38. As gainfully occupied persons in our urban area are 32.20 per cent of the population, the underemployed may be estimated at 8.48 per cent of the population. This may be compared to 2.59 per cent which is the estimate of the unemployed persons. The underemployed are thus 3.28 times as many as the unemployed.

4.39. A more detailed breakdown giving for each industry separately the proportion of the underemployed and the fully employed is given in Table 34 below.

TABLE 34(23, 24) : PERCENTAGE DISTRIBUTION OF UNDEREMPLOYED AND FULLY EMPLOYED PERSONS IN EACH INDUSTRY

industry	under-employed persons	fully employed persons	not recorded	total
(1)	(2)	(3)	(4)	(5)
1. agriculture, animal husbandry, forestry, fishing and hunting	28.09	71.48	0.43	100.00
2. mining and quarrying	11.76	88.24	0.00	100.00
3. manufactures	33.29	65.90	0.81	100.00
4. construction	65.17	34.33	0.50	100.00
5. electricity, water, gas and sanitary services	12.21	87.79	0.00	100.00
6. distributive services	18.95	80.71	0.34	100.00
7. transport, storage and communication	27.57	72.13	0.30	100.00
8. other services not covered above	17.57	82.29	0.14	100.00
9. total	26.32	72.93	0.75	100.00

4.40. Underemployment is most widespread in 'construction' in which industry, almost two-thirds are underemployed. 'Manufactures' comes next with almost one-third underemployed. One would think that in 'agriculture' underemployment is very acute, but here it is about average. This may be due to the fact that agricultural activities in urban areas are of a character quite different from similar activities in rural areas, and are calculated to give fuller and more regular employment. Animal husbandry (dairying, poultry etc.), for example, may play an important role in urban areas and explain the lesser extent of underemployment.

4.41. *Reasons of underemployment.* We have seen that more than 26 per cent of the gainfully occupied persons are underemployed. Various reasons such as disability to accept work on account of illness, fall in production owing to lack of materials and equipments, fall in demand, slack season etc., have been adduced to account for this underemployment. The following Table shows at a glance the relative importance of these various causes of underemployment.

TABLE 35(25b) : PERCENTAGE DISTRIBUTION OF UNDEREMPLOYED PERSONS BY INTENSITY OF EMPLOYMENT AND REASONS FOR UNDEREMPLOYMENT

reasons for underemployment	intensity of employment				
	less than quarter	quarter	half	three quarters	total
(1)	(2)	(3)	(4)	(5)	(6)
1. disability to accept work for illness	27.00	13.01	12.50	20.47	18.26
2. domestic reasons	3.75	4.88	3.99	10.04	5.83
3. fall in production due to lack of materials and equipments	1.75	2.85	2.78	5.71	3.40
4. fall in demand	16.50	33.73	31.60	24.81	26.43
5. slack season	27.25	25.20	20.83	12.40	20.46
6. industrial disputes	0.00	0.00	0.00	0.39	0.12
7. other reasons	22.50	20.33	28.30	26.18	25.21
8. not recorded	1.25	0.00	0.00	0.00	0.29
9. total	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(400)	(246)	(576)	(508)	(1,730)

4.42. In almost half the cases (46.89 per cent), lack of demand is the reason for underemployment. Disability to work on account of illness is another major cause contributing to 18.26 per cent of underemployment. Domestic reasons have been responsible for 5.83 per cent, lack of materials and equipment have added to difficulties of production, resulting in underemployment of 3.40 per cent. There are other reasons of a diverse kind which have contributed to the remaining 25.62 per cent.

4.43. *Income level of the underemployed.* The following Table shows what percentage of the underemployed are included in different income groups. Similar information is given for the fully employed and for both the categories combined. The last three columns give percentages cumulated from below and give at any income level the percentage having an income not lower than the minimum of this income group.

TABLE 36(24) : PERCENTAGE DISTRIBUTION OF UNDEREMPLOYED AND FULLY EMPLOYED PERSONS BY INCOME[1]

income group (Rs. per month)	under-employed	fully employed	total	cumulated percentage from below[1]		
				under-employed	fully employed	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. up to 25	60.05	26.24	35.10	100.00	100.00	100.00
2. 26— 50	23.06	26.95	25.92	39.95	73.76	64.90
3. 51—100	12.30	29.94	25.30	16.89	46.81	38.98
4. 101—150	3.16	9.66	7.96	4.59	16.87	13.68
5. 151—300	1.29	5.46	4.39	1.43	7.21	5.72
6. 301—500	0.07	0.98	0.73	0.14	1.75	1.33
7. above 500	0.07	0.77	0.60	0.07	0.77	0.60
8. total	100.00	100.00	100.00	100.00	100.00	100.00

4.44. We find that only 4.59 per cent of the underemployed have monthly incomes exceeding Rs. 100 and only 16.89 per cent have incomes exceeding Rs. 50. As many as 60.05 per cent have incomes below Rs. 25.

4.45. As is only to be expected, the fully employed are definitely better off in this respect. A little more than half (53.19 per cent) are still having incomes of less than Rs. 50 but this compares favourably with the more than four-fifths (83.11 per cent) of the underemployed included in this income group. The favourable position will be evident primarily from the lower level of income where the percentage falls from 60.05 per cent in the case of the underemployed to 26.24 per cent in the case of the fully employed. For the fully employed the proportion in the higher income groups increases but income above Rs. 300 is reached by only 1.75 per cent and even that of Rs. 150 by not more than 7.21 per cent.

4.46. *Income expectation of the unemployed.* The income expectation of the unemployed whether seeking employment for the first time or not for the first time is seen to follow practically identical pattern and it is not necessary therefore to treat the two sub-groups separately. The following Table gives a comparison between the incomes of the underemployed and the fully employed on the one hand and the income expectations of the unemployed on the other.

---

[1] The 'not recorded' percentage in Table 24 of Statistical Tables has been proportionately distributed in different income groups.

TABLE 37(27) : INCOME EXPECTATIONS OF THE UNEMPLOYED COMPARED WITH THE ACTUAL INCOMES OF THE EMPLOYED[1]

range of income actual/expected (Rs. per month)	percentage in each class[1]		
	unemployed	under-employed	fully employed
(1)	(2)	(3)	(4)
1. up to 25	8.84	60.05	26.24
2. 26— 50	30.65	23.06	26.95
3. 51—100	45.58	12.30	29.94
4. 101—150	12.97	3.16	9.66
5. above 150	1.96	1.43	7.21
6. total	100.00	100.00	100.00

4.47. An income of 'up to Rs. 100' per month will meet the expectation of 85.07 per cent of the unemployed and nearly half of them will be content with an income half as much. Another 12.97 per cent more do not expect anything more than an income of Rs. 100–150 and barely 2 per cent expect a salary exceeding Rs. 150.

4.48. The following Table gives the percentage distribution of the unemployed persons by general education and income expected.

TABLE 38(28b) : PERCENTAGE DISTRIBUTION OF THE UNEMPLOYED OF EACH EDUCATION STANDARD BY RANGE OF INCOME EXPECTED

general education	range of income expected (Rs. per month)						
	up to 25	26–50	51–100	101–150	above 150	not recorded	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. illiterate	22.81	50.00	23.68	0.88	0.00	2.63	100.00
2. literate but below matric	5.97	27.36	51.26	10.06	1.26	4.09	100.00
3. matric	0.00	7.25	59.41	27.54	2.90	2.90	100.00
4. intermediate	0.00	0.00	18.75	68.75	6.25	6.25	100.00
5. graduate and above	0.00	0.00	20.00	50.00	30.00	0.00	100.00
6. total	8.52	29.55	43.94	12.50	1.89	3.60	100.00

It is seen that income expectation of the illiterate centres round the income range Rs. 26–50 per month. Half of them seek this income and the remaining half are almost equally divided between the immediately preceding (up to Rs. 25) and the immediately following (Rs. 51–100) income groups. Half the literates below matric expect Rs. 51–100, a third are content with income of Rs. 26–50 or even less ; and the rest have an income expectation above Rs. 100 per month. The only two important income levels for matriculates are Rs. 51–100 which attracts 59.41 per cent and Rs. 101–150 which attracts 27.54 per cent of the unemployed among them. These incomes are also significant for the intermediates with the important difference that it is the income of Rs. 101–150 which is sought by 68.75 per cent. For the graduates, as might be expected, the expectations are somewhat higher. Half of them expect an income of Rs. 101–150, 30 per cent above Rs. 150 and the rest are satisfied with the middle income range of Rs. 51–100.

---

[1] After distributing the 'not recorded' percentage in Table 27 of Statistical Tables proportionately among the different income groups.

4.49. *Nature of employment sought by the unemployed.* The unemployed in the sample households were asked as to what kind of employment they were seeking. The results of this enquiry are presented in the following Table. Unfortunately, our sample of the unemployed consisted of only 528 persons and this number is inadequate to arrive at valid conclusions in regard to minute classifications. This Table merely illustrates the type of information that could be collected and might be of very great value if a suitably large sample is taken. This is being attempted in the ninth round of the NSS and may be expected to give interesting results.

TABLE 39(30) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY TYPES AND NATURE OF EMPLOYMENT SOUGHT

nature of employment sought	unemployed persons seeking employment: for the first time	unemployed persons seeking employment: not for the first time	unemployed persons seeking employment: total
(1)	(2)	(3)	(4)
as own-account worker			
1. agricultural production	0.00	1.37	0.76
2. animal husbandry	0.84	0.69	0.76
3. manufacture and handicrafts	5.49	8.59	7.20
4. transport service	2.11	2.41	2.27
5. trade	5.49	11.00	8.52
6. professions and financial transac.	2.53	6.53	4.73
7. sub-total (rows 1–6)	16.46	30.59	24.24
as employee			
8. manual work—unskilled	24.04	23.72	23.85
9. manual work—skilled	12.66	15.12	14.02
10. sub-total (rows 8–9)	36.70	38.84	37.87
non-manual work			
11. administrative, executive or clerical work	21.09	11.68	15.90
12. technical and technological	3.80	4.81	4.36
13. medical, surgical, etc.	1.27	1.37	1.33
14. legal, teaching, writing and the like	4.22	3.44	3.79
15. fine arts	1.27	0.00	0.57
16. others	11.39	4.12	7.39
17. sub-total (rows 11–16)	43.04	25.42	33.34
18. sub-total (rows 8–16)	79.74	64.26	71.21
19. not recorded	3.80	5.15	4.55
20. total (rows 7, 18, 19)	100.00	100.00	100.00
no. of sample persons	(237)	(291)	(528)

4.50. The broad conclusions that may be drawn are that one-fourth of the unemployed prefer to work on their own account and the rest want jobs as employees. One-third of the latter seek job as unskilled manual workers, one-fifth as skilled manual workers. The rest, wanting non-manual employment, are almost equally divided between those seeking administrative and clerical work and those seeking technical and professional work.

4.51. *Attitude of the unemployed towards registration with employment exchanges.* The unemployed persons were asked whether on the date of the survey they were registered or had applied for registration at one of the employment exchanges, of which there is a net-work all over India. Those who were not registered were asked to state why they had not. The following Table, which summarises the results of this enquiry, shows that in the matter of registration, the two types of the unemployed, those seeking employment for the first time and those not for the first time, display almost an identical pattern.

TABLE 40(32) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY TYPES AND ATTITUDE TO REGISTRATION WITH EMPLOYMENT EXCHANGES

attitude to registration	unemployed persons seeking employment[1]		
	for the first time	not for the first time	total
(1)	(2)	(3)	(4)
1. registered	26.75	27.24	27.02
not registered due to			
2. ignorance	37.55	38.06	37.84
3. remoteness	1.41	0.75	1.04
4. lack of faith	11.74	11.19	11.44
5. indifference	4.69	4.47	4.58
6. ineligibility	2.84	2.24	2.49
7. other reasons	15.02	16.05	15.59
8. sub-total (rows 2–7)	73.25	72.76	72.98
9. not recorded	—	—	—
10. total	100.00	100.00	100.00

4.52. An important conclusion that may be drawn is that only a little more than one-fourth of the unemployed in the urban areas are registered with the employment exchanges. This is in agreement also with the ratio obtained in respect of the unemployed in the Calcutta Employment Survey of the NSS.

4.53. The overall ratio of the registered and the unregistered, however, gives no clue to the large differences observable in this respect between the unemployed having different general education standard. The differential pattern is brought out in the following Table.

---

[1] Percentages have been calculated after distributing the 'not recorded' percentages of Table 32 of Statistical Tables in the different rows proportionately.

TABLE 41(33) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY GENERAL EDUCATION AND ATTITUDE TO REGISTRATION

attitude to registration	general education						
	illiterate	literate but below matric	matric	inter-mediate	graduate and above	not recorded	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. registered	14.04	24.53	39.13	43.75	20.00	0.00	24.62
not registered due to							
2. ignorance	56.14	34.90	5.80	12.50	10.00	0.00	34.47
3. remoteness	0.00	0.63	2.90	6.25	0.00	0.00	0.95
4. lack of faith	5.26	10.38	15.94	6.25	40.00	0.00	10.42
5. indifference	3.51	3.77	7.25	0.00	10.00	0.00	4.17
6. ineligibility	3.51	2.52	0.00	0.00	0.00	0.00	2.27
7. other reasons	11.40	14.15	15.94	25.00	10.00	100.00	14.20
8. sub-total (rows 2–7)	79.82	66.35	47.83	50.00	70.00	100.00	66.48
9. not recorded	6.14	9.12	13.04	6.25	10.00	0.00	8.90
10. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(114)	(318)	(69)	(16)	(10)	(1)	(528)

4.54. To ensure better comparison, we may distribute the 'not recorded' under each column proportionately among the different rows, and present the result as below.

TABLE 42 : PROPORTION OF THE REGISTERED TO THE TOTAL UNEMPLOYED OF EACH EDUCATION STANDARD

general education	proportion of registered to unemployed	ratio of total unemployed to the no. registered
(1)	(2)	(3)
1. illiterate	0.15	6.67
2. literate but below matric	0.27	3.70
3. matric	0.45	2.22
4. intermediate	0.47	2.13
5. graduate and above	0.22	4.54
6. total	0.27	3.70

4.55. There are proportionately far more 'matriculates' and 'intermediates' and, though to a somewhat lesser extent, the 'literates below matric', represented among the registered, than illiterates and graduates. The pattern of unemployment as revealed by registration data alone would, therefore, give an altogether distorted picture of the relative proportion of persons with different general education standard among the unemployed as a whole. Column 3 of the Table above gives the factors by which the number registered would have to be multiplied, after the registered have been classified by education standard, to arrive at estimates more correctly representing the composition of the unemployed population.

4.56. The relative importance of the different reasons which influence the number of the unemployed registering themselves is different for different education standards.

TABLE 43(33) : RELATIVE IMPORTANCE OF DIFFERENT REASONS FOR NON-REGISTRATION (Base : Non-registered unemployed persons of each education standard)

reasons for non-registration	general education				
	illiterate	literate but below matric	matric and intermediates	graduates	total
(1)	(2)	(3)	(4)	(5)	(6)
1. ignorance	70.33	52.60	14.63	14.29	51.85
2. remoteness	0.00	0.95	7.32	0.00	1.43
3. lack of faith	6.59	15.64	29.27	57.13	15.67
4. indifference	4.40	5.68	12.20	14.29	6 27
5. ineligibility	4.40	3.80	0.00	0.00	3.42
6. other reasons	14.28	21.33	36.58	14.29	21.36
7. total	100.00	100 00	100.00	100.00	100.00

4.57. The main reason for the illiterates not registering themselves in the Exchanges is their 'ignorance' about them. This holds good, though to a much lesser extent, of the 'literate below matric' group of whom more than half are not aware of the existence of the Exchanges. 'Ignorance' ceases to play an important role for the educated group of whom all but one in seven appear to know about the Exchanges. This knowledge unfortunately seems to have brought about only disillusionment which is shown by the increasingly high proportion of those ascribing their non-registration to 'lack of faith' in the value of registration itself, as education standard goes up. Starting from 6.59 per cent of illiterates displaying this 'lack of faith', one comes across 15.64 per cent of them among 'literates below matric', 29.27 per cent among the matriculates and intermediates and as many as 57.13 per cent among the graduates. The trend of 'indifference' as a contributory cause for non-registration follows a similar course.

4.58. On the whole, more than half who have not registered have done so because of their ignorance of the facilities for it. Of the rest, it is worth noting, almost a third have no faith in the efficacy of registration as a means to secure employment.

4.59. *Estimate of unemployment.* If the assumption can be made that the proportion of the registered to the total unemployed remains fairly stable over time, the registration figures of the Employment Exchanges may be made to yield information regarding the total volume of unemployment as well as the trend in its rise and fall, by the use of the information given in the preceding paragraphs.

4.60. For example, the number registered at the Exchanges in September 1953, the period of the present enquiry, was 512,000. The unemployed in the urban areas would be, using our ratio, 3.70 times as many as the number registered, that is about 1.89 million. Of the same order is the estimate obtained directly from our survey, assuming that the level of unemployment is the same in towns of population below 50,000 as in towns of larger size and after allowing for the higher level of unemployment in the big cities like Calcutta, which it does not cover. We have found that 2.59 per cent of the population is unemployed in the towns covered by our survey ; in Calcutta the unemployed represent 7.10 per cent of the population, according to a sample survey conducted by the NSS in Calcutta towards the end of 1953. On a rough estimate based on Census 1951 population[1] after adjusting for increase in population at the average annual rate of 5 per cent for the four big cities and 3 per cent for other urban areas[1] during the period of 30 months between the 1951 Census and the present survey, it may be assumed that a total of 67.99 million persons were living in the urban areas of India at the time of the survey of which 10.54 million were living in the four big cities and 57.45 million in the other urban areas. Applying the percentage proportion of the unemployed for those two types of urban areas, we arrive at an estimate of 0.75 million unemployed in the four big cities and 1.49 million in other urban areas giving an overall estimate of 2.24 million, which is of the same order as the rough estimate of 1.89 million based on employment exchange data.

4.61. This is, however, an estimate of the unemployed on the basis of a stringent definition of unemployment. We have seen (para 4.39) that more than three times as many are the underemployed. Underemployment, however, is of varying intensity. Those who have work for a quarter of the time or less are much closer to the unemployed than those who have work for three quarters of the time or even half the time. A reasonable assumption may be to treat those with intensity of employment 'quarter or less' as coming within the purview of an enlarged definition of unemployment. They represent 3.17 per cent of the population. Adding to this number the unemployed who are 2.59 per cent of the population, we may roughly estimate the level of unemployment and severe underemployment at 5.76 per cent of the population.

---

[1] Decennial increase in population of the four big cities [(1) Calcutta and Howrah; (2) Greater Bombay; (3) Madras and (4) Delhi and New Delhi] was 51 per cent during 1941–51 and 50 per cent during 1931–41; corresponding figures for the other urban areas are 34 per cent and 28 per cent respectively (Census of India 1951, Vol. I, Part II-A, p. 64).

4.62. Generalising on this basis for the urban population of 57.45 million (not including the population of 10.5 million in the four big cities, Calcutta, Bombay, Madras and Delhi) we arrive at an estimate of 57.45 × 0.0317 million or roughly 1.82 million for the number of severely underemployed in these areas.

4.63. If it is assumed that in the four big cities also the relative proportion of the severely underemployed (having intensity of employment 'quarter or less') and the totally unemployed is, as in the urban areas, covered by the survey, we may make the rough estimate of 0.92 million for the number of severely underemployed in these cities, after making assumptions as in para 4.60. This is likely to be an overestimate in view of the different (and more liberal) definition of unemployment adopted in the NSS Calcutta survey, as discussed earlier in para 1.19.

4.64. For the urban area as a whole, therefore, we arrive at an estimate of 2.24 million unemployed together with roughly another 2.74 million severely underemployed. It is necessary, however, to emphasise the provisional character of these estimates, especially in relation to underemployment which is based purely on our subjective evaluation of the persons investigated.

4.65. An estimate may also be made of the educated unemployed in the urban areas on the basis of the statistics of Employment Exchanges relating to the number of matriculates, intermediates and graduates on the live registers in September 1953 by applying to them the ratios of registered to the total unemployed obtained in our survey. The results are presented in the following Table.

TABLE 44 : ESTIMATE OF THE EDUCATED UNEMPLOYED BY USING EMPLOYMENT EXCHANGE DATA—SEPTEMBER 1953

education standard	no. on live registers in September 1953 (in thousands)	ratio of registered to unemployed	estimated no. of unemployed (in thousands)
(1)	(2)	(3)	(4)
1. matriculates	125.8	0.45	279.6
2. intermediates	16.9	0.47	35.9
3. graduates and post-graduates	19.6	0.22	89.1
4. total	162.3	—	404.6

This Table gives an estimate of 4.05 lakhs for the number of educated unemployed.

4.66. Another estimate of their number may be made by using the information of Table 30 which shows that 18 per cent of the total number of the unemployed are matriculates, intermediates and graduates. The estimated number of unemployed, as discussed above, is 22.4 lakhs and 18 per cent of this number is 4.03 lakhs, which is seen to be in good agreement with the estimate arrived at independently by the first method.

NATIONAL SAMPLE SURVEY : URBAN UNEMPLOYMENT

STATISTICAL TABLES

PAGE

CLASSIFID LIST OF TABLES 71

TABLES ... ...

CLASSIFIED LIST OF TABLES

TABLE 1 : Distribution of sample towns, sample blocks and sample households 74

DESCRIPTION OF GENERAL POPULATION

TABLE 2 : Percentage distribution of population by sex and age-groups
  a) base : all persons 75
  b) base : persons of each sex 75
  c) base : persons of each age-group 75

TABLE 3 : Percentage distribution of population by sex and economic status
  a) base : all persons 76
  b) base : persons of each sex 76
  c) base : persons of each economic status 76

TABLE 4 : Percentage distribution of population by age-groups and economic status
  a) base : persons of all ages 77
  b) base : persons of each age-group 77
  c) base : persons of each economic status 78

TABLE 5 : Percentage distribution of males by age-groups and economic status
  a) base : males of each age-group 79
  b) base : males of each economic status 79

TABLE 6 : Percentage distribution of females by age-groups and economic status
  a) base : females of each age-group 80
  b) base : females of each economic status 80

TABLE 7 : Percentage distribution of population by household expenditure classes and economic status
  a) base : persons of each expenditure class 81
  b) base : persons of each economic status 81

TABLE 8 : Percentage distribution of population by sex, general education and technical or professional qualification
  a) base : all persons 82
  b) base ; persons of each sex 83
  c) base : persons of each educational category 84

TABLE 9 : Percentage distribution of population by sex and industrial status
  a) base : all persons 85
  b) base : persons of each sex 86
  c) base : persons of each industrial status 87

TABLE 10 : Percentage distribution of population by age-groups and industrial status
  a) base : persons of each age-group 88
  b) base : persons of each industrial status 89

Page

Table 11 : Percentage distribution of males by age-groups and industrial status
a) base : males of each age-group 90
b) base : males of each industrial status 91

Table 12 : Percentage distribution of females by age-groups and industrial status
a) base : females of each age-group 92
b) base : females of each industrial status 93

Table 13 : Percentage distribution of population by industrial status and general education
a) base : persons of each general education standard 94
b) base : persons of each industrial status 95

Table 14 : Percentage distribution of population by household expenditure classes and industrial status
a) base : persons of each household expenditure class 96
b) base : persons of each industrial status 97

Labour force

Table 15 : Percentage distribution of persons in the labour force by industrial status and types of migrants and non-migrants
a) base : persons of each industrial status 98
b) base : persons in the labour force belonging to each type of migrants or non-migrants 99

Table 16 : Percentage distribution of migrants in the labour force by industrial status and area from where migrated
base : migrants of each industrial status 100

Table 17 : Percentage distribution of migrants in the labour force by area from where migrated and reasons for migration
a) base : migrants in the labour force having each type of reason for migration .. 101
b) base : migrants in the labour force from each area of migration 102

Table 18 : Percentage distribution of migrants in the labour force by area from where migrated and ranges of duration of migration
a) base : migrants in the labour force having each range of duration of migration .. 103
b) base : migrants in the labour force from each area of migration 104

Table 19 : Percentage distribution of migrants in the labour force by reasons for migration and ranges of duration of migration
a) base : migrants in the labour force having each range of duration of migration .. 105
b) base : migrants in the labour force having each type of reason for migration .. 106

Gainfully occupied persons

Table 20 : Percentage distribution of gainfully occupied persons by industry and industrial status
a) base : gainfully occupied persons belonging to each industry 107
b) base : persons of each industrial status 107

Table 21 : Percentage distribution of gainfully occupied persons by industry and income groups
a) base : gainfully occupied persons of each income group 108
b) base : gainfully occupied persons belonging to each industry 109

PAGE

TABLE 22 : Percentage distribution of employees by general education and types of employer
a) base : employees having each type of employer 110
b) base : employees of each general education standard 110

TABLE 23 : Percentage distribution of gainfully occupied persons in each industry by intensity of employment 111

TABLE 24 : Percentage distribution of gainfully occupied persons by intensity of employment and income 112

TABLE 25 : Percentage distribution of underemployed gainfully occupied persons by intensity of employment and reasons for underemployment 113

UNEMPLOYED

TABLE 26 : Percentage distribution of unemployed persons by types and education 114

TABLE 27 : Percentage distribution of unemployed persons by types and income expected .. 115

TABLE 28 : Percentage distribution of unemployed persons by general education and income expected 116

TABLE 29 : Percentage distribution of unemployed persons by nature of employment sought and general education
a) base : unemployed persons with each type of employment sought 117
b) base : unemployed persons of each general education standard 118

TABLE 30 : Percentage distribution of unemployed persons by types and nature of employment sought 119

TABLE 31 : Percentage distribution of unemployed persons by nature of employment sought and income expected
a) base : unemployed persons with each type of employment sought 120
b) base : unemployed persons with each range of income expected 121

TABLE 32 : Percentage distribution of unemployed persons by types and attitude to registration 122

TABLE 33 : Percentage distribution of unemployed persons by general education and attitude to registration 123

TABLE (1) : DISTRIBUTION OF SAMPLE TOWNS, SAMPLE BLOCKS AND SAMPLE HOUSEHOLDS

population zone	population size-class						total		
	towns with population of not less than 100,000			towns with population between 50,000 and 100,000					
	no. of sample towns	no. of sample blocks	no. of sample households	no. of sample towns	no. of sample blocks	no. of sample households	no. of sample towns	no. of sample blocks	no. of sample households
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. North and North-West India	5	50	1,012	2	16	342	7	66	1,354
2. East India	2	20	395	2	16	384	4	36	779
3. Central India	3	30	589	2	16	288	5	46	877
4. South and West India	4	40	804	3	24	632	7	64	1,436
5. all India	14	140	2,800	9	72	1,646	23	212	4,446

TABLE (2) : PERCENTAGE DISTRIBUTION OF POPULATION BY SEX AND AGE-GROUPS

		(a) base : all persons			(b) base : persons of each sex			(c) base : persons of each age-group			
	age-group (in years)	male	female	total	male	female	total	male	female	total	no. of sample persons
	(1)	(2)	(3)	(4)	(2)	(3)	(4)	(2)	(3)	(4)	(5)
1.	0— 6	9.88	9.73	19.61	19.14	20.12	19.61	50.40	49.60	100.00	(4,002)
2.	7—15	10.65	10.30	20.95	20.62	21.29	20.95	50.84	49.16	100.00	(4,274)
3.	16—17	1.95	1.66	3.61	3.77	3.43	3.61	54.01	45.99	100.00	(735)
4.	18—21	4.23	4.23	8.46	8.20	8.75	8.46	50.03	49.97	100.00	(1,727)
5.	22—26	5.12	4.82	9.94	9.93	9.96	9.94	51.55	48.45	100.00	(2,029)
6.	27—36	7.86	6.54	14.40	15.21	13.53	14.40	54.56	45.44	100.00	(2,938)
7.	37—46	5.35	4.56	9.91	10.34	9.43	9.91	53.93	46.07	100.00	(2,021)
8.	47—56	3.60	3.27	6.87	6.97	6.77	6.87	52.39	47.61	100.00	(1,403)
9.	57—61	1.29	1.37	2.66	2.51	2.83	2.66	48.62	51.38	100.00	(543)
10.	62 & above	1.71	1.88	3.59	3.31	3.89	3.59	47.61	52.39	100.00	(733)
11.	total	51.64	48.36	100.00	100.00	100.00	100.00	51.64	48.36	100.00	(20,405)
	number of sample persons				(10,538)	(9,867)	(20,405)				

number of sample towns=23 ; number of sample blocks=212; number of sample households=4446 ; average household size=4.59

TABLE (3) : PERCENTAGE DISTRIBUTION OF POPULATION BY SEX AND ECONOMIC STATUS

	(a) base : all persons			(b) base : persons of each sex			(c) base : persons of each economic status		
economic status	male	female	total	male	female	total	male	female	total
(1)	(2)	(3)	(4)	(2)	(3)	(4)	(2)	(3)	(4)
1. earners	23.37	2.89	26.26	45.26	5.98	26.26	89.01	10.99	100.00
2. earning dependents	4.35	3.65	8.00	8.41	7.54	8.00	54.35	45.65	100.00
3. non-earning dependents	23.81	41.75	65.56	46.11	86.33	65.56	36.33	63.67	100.00
4. not recorded	0.11	0.07	0.18	0.22	0.15	0.18	60.53	39.47	100.00
5. total	51.64	48.36	100.00	100.00	100.00	100.00	51.64	48.36	100.00
no. of sample persons	(10,538)	(9,867)	(20,405)	(10,538)	(9,867)	(20,405)	(10,538)	(9,867)	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446 ;

TABLE (4) : PERCENTAGE DISTRIBUTION OF POPULATION BY AGE-GROUPS AND ECONOMIC STATUS

(a) base : persons of all ages

economic status	age-groups											no. of sample persons
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above	total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. earners	0.01	0.44	0.39	2.03	4.14	7.88	5.71	3.65	1.04	0.97	26.26	(5,358)
2. earning dependents	0.00	1.42	0.64	1.25	1.31	1.29	0.73	0.59	0.34	0.43	8.00	(1,632)
3. non-earning dependents	19.57	19.04	2.56	5.17	4.48	5.21	3.46	2.62	1.27	2.18	65.56	(13,377)
4. not recorded	0.03	0.05	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.18	(38)
5. total	19.61	20.95	3.60	8.46	9.94	14.40	9.91	6.88	2.66	3.59	100.00	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

(b) base : persons of each age-group

economic status	age-groups										total
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(8)	(10)	(11)	(12)
1. earners	0.05	2.08	10.75	24.03	41.60	54.77	57.64	53.03	39.23	27.01	26.26
2. earning dependents	0.02	6.79	17.82	14.71	13.16	8.91	7.33	8.62	12.89	12.01	8.00
3. non-earning dependents	99.78	90.88	71.02	61.09	45.10	36.15	34.98	38.06	47.70	60.85	65.56
4. not recorded	0.15	0.25	0.41	0.17	0.14	0.17	0.05	0.29	0.18	0.13	0.18
5. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
number of sample persons	(4 002)	(4,274)	(735)	(1,727)	(2,029)	(2,938)	(2,021)	(1,403)	(543)	(733)	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (4) : PERCENTAGE DISTRIBUTION OF POPULATION BY AGE-GROUPS AND ECONOMIC STATUS

(c) base : persons of each economic status

economic status	age-groups										total	no. of sample persons
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. earners	0.04	1.66	1.47	7.75	15.75	30.02	21.74	13.89	3.98	3.70	100.00	(5,358)
2. earning dependents	0.06	17.78	8.03	15.56	16.36	16.05	9.07	7.41	4.29	5.39	100.00	(1,632)
3. non-earning dependents	29.85	29.03	3.90	7.89	6.84	7.94	5.29	3.99	1.94	3.33	100.00	(13,377)
4. not recorded	15.79	28.96	7.89	7.89	7.89	13.16	2.63	10.53	2.63	2.63	100.00	(38)
5. total	19.61	20.95	3.60	8.46	9.94	14.40	9.91	6.88	2.66	3.59	100.00	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (5) : PERCENTAGE DISTRIBUTION OF MALES BY AGE-GROUPS AND ECONOMIC STATUS

(a) base : males of each age-group

economic status	age-groups										total
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1. earners	0.05	3.08	18.14	43.87	74.95	90.33	93.67	88.56	67.80	47.85	45.26
2. earning dependents	0.00	9.90	25.93	20.95	14.92	5.68	2.66	3.95	10.61	15.76	8.41
3. non-earning dependents	99.80	86.74	55.43	34.95	9.94	3.68	3.58	7.35	21.21	36.39	46.11
4. not recorded	0.15	0.28	0.50	0.23	0.19	0.31	0.09	0.14	0.38	0.00	0.22
5. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
number of sample persons	(2,017)	(2,173)	(397)	(864)	(1,046)	(1,603)	(1,090)	(735)	(264)	(349)	(10,538)

(b) base : males of each economic status

economic status	age-groups										total	no. of sample persons
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. earners	0.02	1.40	1.51	7.95	16.44	30.37	21.41	13.65	3.75	3.50	100.00	(4,769)
2. earning dependents	0.00	24.23	11.61	20.41	17.59	10.26	3.27	3.27	3.16	6.20	100.00	(887)
3. non-earning dependents	41.44	38.79	4.53	6.22	2.14	1.21	0.80	1.11	1.15	2.61	100.00	(4,859)
4. not recorded	13.03	26.08	8.70	8.70	8.70	21.74	4.35	4.35	4.35	0.00	100.00	(23)
5. total	19.14	20.62	3.77	8.20	9.93	15.21	10.34	6.97	2.51	3.31	100.00	(10,538)

number of sample towns=23; number of sample blocks=212; number of sample households=4,446

TABLE (6) : PERCENTAGE DISTRIBUTION OF FEMALES BY AGE-GROUPS AND ECONOMIC STATUS

(a) base : females of each age-group

economic status	age-groups										total
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1. earners	0.05	1.05	2.07	4.17	6.10	12.06	15.47	13.92	12.19	8.07	5.98
2. earning dependents	0.05	3.56	8.28	8.46	11.30	12.81	12.78	13.77	15.05	8.59	7.54
3. non-earning dependents	99.75	95.15	89.35	87.25	82.50	75.13	71.75	71.86	72.76	83.08	86.33
4. not recorded	0.15	0.24	0.30	0.12	0.01	0.00	0.00	0.45	0.00	0.26	0.15
5. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(1,985)	(2,101)	(338)	(863)	(983)	(1,335)	(931)	(668)	(279)	(384)	(9,867)

(b) base : females of each economic status

economic status	age-groups										total	no. of sample persons
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. earners	0.17	3.74	1.19	6.11	10.19	27.33	24.45	15.79	5.77	5.26	100.00	(589)
2. earning dependents	0.13	10.07	3.76	9.80	14.90	22.95	15.97	12.35	5.64	4.43	100.00	(745)
3. non-earning dependents	23.24	23.47	3.55	8.84	9.52	11.78	7.84	5.64	2.38	3.74	100.00	(8,518)
5. not recorded	20.00	33.32	6.67	6.67	6.67	0.00	0.00	20.00	0.00	6.67	100.00	(15)
total	20.12	21.29	3.43	8.75	9.96	13.53	9.43	6.77	2.83	3.89	100.00	(9,867)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (7) : PERCENTAGE DISTRIBUTION OF POPULATION BY HOUSEHOLD EXPENDITURE CLASSES AND ECONOMIC STATUS

(a) base : persons of each expenditure class

economic status	expenditure classes (in rupees per month)								total	no. of sample persons
	upto 50	51–100	101–200	201–300	301–500	501–1000	above 1000	not recorded		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1. earners	33.73	25.51	22.29	20.55	21.25	31.16	32.08	30.25	26.26	(5,358)
2. earning dependents	9.69	8.82	6.72	5.64	5.12	1.40	10.37	4.94	8.00	(1,632)
3. non-earning dependents	56.16	65.60	70.83	73.81	73.63	66.04	57.55	64.19	65.56	(13,377)
4. not recorded	0.42	0.07	0.16	0.00	0.00	1.40	0.00	0.62	0.18	(38)
5. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	
no. of sample persons	(4,954)	(7,375)	(5,136)	(1,596)	(861)	(215)	(106)	(162)	(20,405)	

(b) base : persons of each economic status

1. earners	31.19	35.11	21.37	6.12	3.42	1.25	0.63	0.91	100.00	(5,358)
2. earning dependents	29.41	39.90	21.14	5.51	2.70	0.18	0.67	0.49	100.00	(1,632)
3. non-earning dependents	20.80	36.15	27.20	8.81	4.74	1.06	0.46	0.78	100.00	(13,377)
4. not recorded	55.27	13.16	21.05	0.00	0.00	7.89	0.00	2.63	100.00	(38)
5. total	24.28	36.14	25.17	7.82	4.22	1.05	0.52	0.80	100.00	(20,405)

number of sample towns=23 ; number of sample blocks=212; number of sample households=4,446

TABLE (8) : PERCENTAGE DISTRIBUTION OF POPULATION BY SEX, GENERAL EDUCATION AND TECHNICAL OR PROFESSIONAL QUALIFICATION

(a) base : all persons

general education	technical or professional qualification[1]	male	female	total
(1)	(2)	(3)	(4)	(5)
1. illiterate	A	19.33	32.22	51.55
	B	4.40	1.41	5.81
	sub-total	23.73	33.63	57.36
2. literate but below matric	A	17.13	12.88	30.01
	B	6.45	1.10	7.55
	C	0.24	0.06	0.30
	sub-total	23.82	14.04	37.86
3. matric	A	1.51	0.33	1.84
	B	0.73	0.08	0.81
	C	0.32	0.02	0.34
	D	0.17	0.01	0.18
	sub-total	2.73	0.44	3.17
4. intermediate	A	0.36	0.04	0.40
	B	0.16	0.02	0.18
	C	0.09	0.00	0.09
	D	0.08	0.00	0.08
	sub-total	0.69	0.06	0.75
5. graduate and above	A	0.21	0.03	0.24
	B	0.11	0.03	0.14
	C	0.04	0.00	0.04
	D	0.22	0.00	0.22
	sub-total	0.58	0.06	0.64
6. not recorded		0.09	0.13	0.22
7. total	A	38.54	45.50	84.04
	B	11.85	2.64	14.49
	C	0.69	0.08	0.77
	D	0.47	0.01	0.48
	not recorded	0.09	0.13	0.22
	sub-total	51.64	48.36	100.00
no. of sample persons		(10,538)	(9,867)	(20,405)

number of sample towns = 23 ; sample blocks = 212 ; sample households = 4,446

[1] Technical or professional qualification :

A. no technical or professional qualification ;
B. professional or technical skill without any degree or equivalent diploma and not covered by C;
C. certificate or diploma in typing, stenography, nursing, midwifery, or pharmacy;
D. degree or equivalent diploma in any technology or profession.

TABLE (8) : PERCENTAGE DISTRIBUTION OF POPULATION BY SEX, GENERAL EDUCATION AND TECHNICAL OR PROFESSIONAL QUALIFICATION

(b) base : persons of each sex

general education	technical or professional qualification[1]	male	female	total
(1)	(2)	(3)	(4)	(5)
1. illiterate	A	37.44	66.62	51.55
	B	8.51	2.92	5.81
	sub-total	45.95	69.54	57.36
2. literate but below matric	A	33.18	26.62	30.01
	B	12.49	2.28	7.55
	C	0.47	0.12	0.30
	sub-total	46.14	29.02	37.86
3. matric	A	2.93	0.67	1.84
	B	1.41	0.17	0.81
	C	0.63	0.04	0.34
	D	0.32	0.03	0.18
	sub-total	5.29	0.91	3.17
4. intermediate	A	0.71	0.09	0.40
	B	0.31	0.05	0.18
	C	0.17	0.00	0.09
	D	0.14	0.00	0.08
	sub-total	1.33	0.14	0.75
5. graduate and above	A	0.41	0.05	0.24
	B	0.21	0.06	0.14
	C	0.09	0.00	0.04
	D	0.42	0.00	0.22
	sub-total	1.13	0.11	0.64
6. not recorded		0.16	0.28	0.22
7. total	A	74.67	94.05	84.04
	B	22.93	5.48	14.49
	C	1.36	0.16	0.77
	D	0.88	0.03	0.48
	not recorded	0.16	0.28	0.22
	sub-total	100.00	100.00	100.00
no. of sample persons		(10,538)	(9,867)	(20,405)

number of sample towns = 23 ; sample blocks = 212 ; sample households = 4,446

[1] Technical or professional qualification :

A. no technical or professional qualification;
B. professional or technical skill without any degree or equivalent diploma and not covered by C
C. certificate or diploma in typing, stenography, nursing, midwifery, or pharmacy;
D. degree or equivalent diploma in any technology or profession.

TABLE (8) : PERCENTAGE DISTRIBUTION OF POPULATION BY SEX, GENERAL EDUCATION AND TECHNICAL OR PROFESSIONAL QUALIFICATION

(c) base : persons of each educational category

general education	technical or professional qualification[1]	male	female	total	no. of sample persons
(1)	(2)	(3)	(4)	(5)	(6)
1. illiterate	A	37.50	62.50	100.00	(10,519)
	B	75.70	24.30	100.00	(1,185)
	sub-total	41.37	58.63	100.00	(11,704)
2. literate but below matric	A	57.10	42.90	100.00	(6,122)
	B	85.40	14.60	100.00	(1,541)
	C	80.65	19.35	100.00	(62)
	sub-total	62.94	37.06	100.00	(7,725)
3. matric	A	82.40	17.60	100.00	(375)
	B	89.70	10.30	100.00	(165)
	C	94.29	5.71	100.00	(70)
	D	91.89	8.11	100.00	(37)
	sub-total	86.09	13.91	100.00	(647)
4. intermediate	A	89.16	10.84	100.00	(83)
	B	86.84	13.16	100.00	(38)
	C	100.00	0.00	100.00	(18)
	D	100.00	0.00	100.00	(15)
	sub-total	90.91	9.09	100.00	(154)
5. graduate and above	A	89.58	10.42	100.00	(48)
	B	78.57	21.43	100.00	(28)
	C	100.00	0.00	100.00	(9)
	D	100.00	0.00	100.00	(45)
	sub-total	91.54	8.46	100.00	(130)
6. not recorded		40.00	60.00	100.00	(45)
7. total	A	45.88	54.12	100.00	(17,147)
	B	81.70	18.30	100.00	(2,957)
	C	89.94	10.06	100.00	(159)
	D	96.91	3.09	100.00	(97)
	not recorded	40.00	60.00	100.00	(45)
	sub-total	51.64	48.36	100.00	(20,405)

number of sample towns=23 ; sample blocks=212 ; sample households=4,446

[1] Technical or professional qualification :

A. no technical or professional qualification;
B. professional or technical skill without any degree or equivalent diploma and not covered by C;
C. certificate or diploma in typing, stenography, nursing, midwifery, or pharmacy;
D. degree or equivalent diploma in any technology or profession.

TABLE (9) : PERCENTAGE DISTRIBUTION OF POPULATION BY SEX AND INDUSTRIAL STATUS

	(a) base : all persons			(b) base : persons of each sex			(c) base :persons of each industrial status			
industrial status	male	female	total	male	female	total	male	female	total	no. of sample persons
(1)	(2)	(3)	(4)	(2)	(3)	(4)	(2)	(3)	(4)	(5)
gainfully occupied persons:										
1. employers	0.43	0.04	0.47	0.83	0.07	0.47	92.63	7.37	100.00	(95)
2. employees	14.56	3.03	17.59	28.18	6.28	17.59	82.73	17.27	100.00	(3,590)
3. own-account workers	9.88	1.52	11.40	19.12	3.15	11.40	86.63	13.37	100.00	(2,326)
4. unpaid household labour	1.65	1.09	2.74	3.20	2.24	2.74	60.39	39.61	100.00	(558)
(i) sub-total :	26.52	5.68	32.20	51.33	11.74	32.20	82.36	17.64	100.00	(6,569)
unemployed persons seeking employment:										
5. for the first time	0.99	0.17	1.16	1.93	0.35	1.16	85.65	14.35	100.00	(237)
6. not for the first time	1.24	0.19	1.43	2.39	0.40	1.43	86.60	13.40	100.00	(291)
(i) sub-total :	2.23	0.36	2.59	4.32	0.75	2.59	86.17	13.83	100.00	(528)
persons not in the labour force:										
7. family members engaged in domestic work only	0.32	21.96	22.28	0.62	45.42	22.28	1.43	98.57	100.00	(4,547)
8. rentiers and pensioners	0.45	0.16	0.61	0.87	0.33	0.61	73.60	26.40	100.00	(125)
9. beggars and persons living on charities and remittances	0.23	0.59	0.82	0.45	1.22	0.82	28.14	71.86	100.00	(167)
10. old and infirm persons, children and students	21.51	19.34	40.95	41.86	39.98	40.95	52.79	47.21	100.00	(8,356)
(iii) sub-total :	22.61	42.05	64.66	43.80	86.95	64.66	34.98	65.02	100.00	(13,195)
11. not recorded	0.28	0.27	0.55	0.55	0.56	0.55	51.33	48.67	100.00	(113)
(iv) total :	51.64	48.36	100.00	100.00	100.00	100.00	51.64	48.36	100.00	(20,405)
number of sample persons	(10,538)	(9,867)	(20,405)	(10,538)	(9,867)	(20,405)				

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (10) : PERCENTAGE DISTRIBUTION OF POPULATION BY AGE-GROUPS AND INDUSTRIAL STATUS

(a) base : persons of each age-group

industrial status	age-groups (in years)										total
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	12)
gainfully occupied persons:											
1. employers	0.00	0.00	0.00	0.40	0.54	0.88	1.48	0.86	0.55	0.82	0.47
2. employees	0.07	4.49	13.88	21.60	30.61	35.94	34.49	27.09	19.34	8.32	17.59
3. own-account workers	0.00	1.76	6.94	8.63	15.62	21.24	24.19	26.09	22.84	17.87	11.40
4. unpaid household labour	0.00	2.41	6.26	6.08	5.62	3.44	1.78	2.42	1.66	1.36	2.74
(i) sub-total :	0.07	8.66	27.08	36.71	52.39	61.50	61.94	56.46	44.39	28.37	32.20
unemployed persons seeking employment:											
5. for the first time	0.00	0.68	5.71	5.85	2.02	0.44	0.40	0.14	0.18	0.00	1.16
6. not for the first time	0.00	0.21	1.77	3.65	3.01	2.25	1.88	2.35	0.74	0.55	1.43
(ii) sub-total :	0.00	0.89	7.48	9.50	5.03	2.69	2.28	2.49	0.92	0.55	2.59
persons not in the labour force:											
7. family members engaged in domestic work only	0.35	3.10	33.47	40.59	39.58	33.19	31.96	27.80	21.36	13.10	22.28
8. rentiers and pensioners	0.00	0.00	0.00	0.11	0.20	0.58	1.19	1.85	3.13	4.77	0.61
9. beggars and persons living on charities and remittances	0.00	0.09	0.27	0.41	0.49	0.85	0.89	2.71	4.42	5.32	0.82
10. old and infirm persons, children and students	98.55	76.56	31.43	12.33	1.97	0.78	1.39	8.48	25.41	47.48	40.95
(iii) sub-total :	98.90	89.75	65.17	53.44	42.24	35.40	35.43	40.84	54.32	70.67	64.66
11. not recorded	1.03	0.70	0.27	0.35	0.34	0.41	0.35	0.21	0.37	0.41	0.55
(iv) total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(4,002)	(4,274)	(735)	(1,727)	(2,029)	(2,938)	(2,021)	(1,403)	(543)	(733)	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (10) : PERCENTAGE DISTRIBUTION OF POPULATION BY AGE-GROUPS AND INDUSTRIAL STATUS

(b) base : persons of each industrial status

industrial status	age-groups (in years)										total	no. of sample persons
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
gainfully occupied persons:												
1. employers	0.00	0.00	0.00	7.37	11.58	27.37	31.58	12.63	3.16	6.31	100.00	(95)
2. employees	0.08	5.35	2.84	10.39	17.30	29.42	19.42	10.58	2.92	1.70	100.00	(3,590)
3. own-account workers	0.00	3.22	2.19	6.41	13.63	26.83	21.02	15.74	5.33	5.63	100.00	(2,326)
4. unpaid household labour	0.00	18.46	8.25	18.82	20.43	18.10	6.45	6.09	1.61	1.79	100.00	(558)
(i) sub-total :	0.05	5.63	3.03	9.65	16.18	27.51	19.06	12.06	3.67	3.16	100.00	(6,569)
unemployed persons seeking employment:												
5. for the first time	0.00	12.24	17.72	42.60	17.30	5.49	3.38	0.85	0.42	0.00	100.00	(237)
6. not for the first time	0.00	3.09	4.47	21.65	20.96	22.68	13.06	11.34	1.37	1.38	100.00	(291)
(ii) sub-total :	0.00	7.20	10.41	31.06	19.32	14.96	8.71	6.63	0.95	0.76	100.00	(528)
persons not in the labour force:												
7. family members engaged in domestic work only	0.31	12.31	5.41	15.42	17.66	21.44	14.21	8.58	2.55	2.11	100.00	(4,547)
8. rentiers and pensioners	0.00	0.00	0.00	1.60	3.20	13.60	19.20	20.80	13.60	28.00	100.00	(125)
9. beggars and persons living on charities and remittances	0.00	2.40	1.20	4.19	5.99	14.97	10.78	22.75	14.37	23.35	100.00	(167)
10. old and infirm persons, children and students	47.20	39.16	2.76	2.55	0.48	0.28	0.34	1.42	1.65	4.16	100.00	(8,356)
(iii) sub-total :	30.00	29.07	3.63	6.99	6.49	7.88	5.43	4.34	2.24	3.93	100.00	(13,195)
11. not recorded	36.28	26.55	1.77	5.31	6.19	10.62	6.20	2.66	1.77	2.65	100.00	(113)
(iv) total	19.61	20.95	3.61	8.46	9.94	14.40	9.91	6.87	2.66	3.59	100.00	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (11) : PERCENTAGE DISTRIBUTION OF MALES BY AGE-GROUPS AND INDUSTRIAL STATUS

(a) base : males of each age-group

industrial status	age-groups (in years)										total
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
gainfully occupied persons:											
1. employers	0.00	0.00	0.00	0.69	0.96	1.50	2.57	1.63	0.76	1.72	0.83
2. employees	0.05	6.35	22.17	36.11	51.72	55.46	52.20	41.50	30.30	13.47	28.18
3. own-account workers	0.00	2.95	11.59	14.93	26.58	34.00	37.61	44.76	39.02	31.81	19.12
4. unpaid household labour	0.00	3.27	7.81	10.07	7.65	2.99	0.92	0.54	—	1.71	3.20
(i) sub-total :	0.05	12.57	41.57	61.80	86.91	93.95	93.30	88.43	70.08	48.71	51.33
unemployed persons seeking employment:											
5. for the first time	0.00	1.15	10.08	10.07	3.63	0.62	0.18	0.14	0.00	0.00	1.93
6. not for the firsttime	0.00	0.41	3.27	6.37	5.16	3.31	2.94	3.95	1.51	0.86	2.39
(ii) sub-total	0.00	1.56	13.35	16.44	8.79	3.93	3.12	4.09	1.51	0.86	4.32
persons not in the labour force:											
7. family members engaged in domestic work only	0.30	1.11	1.76	0.58	0.67	0.31	0.09	0.68	0.75	0.86	0.62
8. rentiers and pensioners	0.00	0.00	0.00	0.12	0.38	0.69	1.47	2.18	4.55	9.17	0.87
9. beggars and persons living on charities and remittances	0.00	0.18	0.50	0.11	0.09	0.19	0.46	0.54	3.41	5.16	0.45
10. old and infirm persons, children and students	98.31	84.12	42.82	20.49	2.87	0.75	1.10	3.81	18.94	34.67	41.86
(iii) sub-total :	98.61	85.41	45.08	21.30	4.01	1.94	3.12	7.21	27.65	49.86	43.80
11. not recorded	1.34	0.46	0.00	0.46	0.29	0.18	0.46	0.27	0.76	0.57	0.55
(iv) total :	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(2 017)	(2,173)	(397)	(864)	(1,046)	(1,603)	(1,090)	(735)	(264)	(349)	(10,538)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (11) : PERCENTAGE DISTRIBUTION OF MALES BY AGE-GROUPS AND INDUSTRIAL STATUS

(b) base : males of each industrial status

industrial status	age-groups (in years)										total	no. of sample persons
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above		
gainfully occupied persons:												
1. employers	0.00	0.00	0.00	6.82	11.36	27.27	31.82	13.64	2.27	6.82	100.00	(88)
2. employees	0.03	4.65	2.96	10.51	18.22	29.93	19.16	10.27	2.69	1.58	100.00	(2,970)
3. own-account workers	0.00	3.17	2.28	6.40	13.80	27.05	20.35	16.33	5.11	5.51	100.00	(2,015)
4. unpaid household labour	0.00	21.07	9.20	25.81	23.74	14.24	2.97	1.19	0.00	1.78	100.00	(337)
(i) sub-total :	0.02	5.05	3.05	9.87	16.80	27.84	18.80	12.01	3.42	3.14	100.00	(5,410)
unemployed persons seeking employment:												
5. for the first time	0.00	12.32	19.70	42.86	18.72	4.93	0.98	0.49	0.00	0.00	100.00	(203)
6. not for the first time	0.00	3.57	5.16	21.82	21.43	21.03	12.70	11.51	1.59	1.19	100.00	(252)
(ii) sub-total :	0.00	7.47	11.65	31.21	20.22	13.85	7.47	6.59	0.88	0.66	100.00	(455)
persons not in the labour force:												
7. family members engaged in domestic work only	9.23	36.92	10.77	7.69	10.77	7.69	1.54	7.69	3.08	4.62	100.00	(65)
8. rentiers and pensioners	0.00	0.00	0.00	1.09	4.35	11.96	17.39	17.39	13.04	34.78	100.00	(92)
9. beggars and persons living on charities & remittances	0.00	8.51	4.26	2.13	2.13	6.38	10.64	8.51	19.14	38.30	100.00	(47)
10. old and infirm persons, children and students	44.96	41.44	3.86	4.01	0.68	0.27	0.27	0.64	1.13	2.74	100.00	(4,411)
(iii) sub-total :	43.10	40.22	3.88	3.99	0.91	0.67	0.74	1.15	1.58	3.76	100.00	(4,615)
11. not recorded	46.55	17.24	0.00	6.90	5.17	5.17	8.62	3.45	3.45	3.45	100.00	(58)
(iv) total :	19.14	20.62	3.77	8.20	9.93	15.21	10.34	6.97	2.51	3.31	100.00	(10,538

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (12) : PERCENTAGE DISTRIBUTION OF FEMALES BY AGE-GROUPS AND INDUSTRIAL STATUS

(a) base : females of each age-group

industrial status	age-groups (in years)										total
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
gainfully occupied persons:											
1. employers	0.00	0.00	0.00	0.12	0.10	0.15	0.21	0.00	0.36	0.00	0.07
2. employees	0.10	2.57	4.14	7.07	8.14	12.51	13.75	11.23	8.96	3.65	6.28
3. own-account workers	0.00	0.52	1.48	2.32	3.97	5.92	8.49	5.54	7.53	5.21	3.15
4. unpaid household labour	0.00	1.52	4.44	2.09	3.46	3.97	2.79	4.49	3.22	1.04	2.24
(i) sub-total :	0.10	4.61	10.06	11.60	15.67	22.55	25.24	21.26	20.07	9.90	11.74
unemployed persons seeking employment:											
5. for the first time	0.00	0.19	0.59	1.62	0.30	0.23	0.64	0.15	0.36	0.00	0.35
6. not for the first time	0.00	0.00	0.00	0.93	0.71	0.97	0.64	0.50	0.00	0.26	0.40
(ii) sub-total :	0.00	0.19	0.59	0.55	1.01	1.20	1.28	0.75	0.36	0.26	0.75
persons not in the labour force:											
7. family members engaged in domestic work only	0.40	25.51	70.71	80.65	80.97	72.66	69.28	57.63	40.86	24.22	45.42
8. rentiers and pensioners	0.00	0.00	0.00	0.11	0.00	0.45	0.86	1.50	1.79	0.78	0.33
9. beggars and persons living on charities and remittances	0.00	0.00	0.00	0.69	0.92	1.65	1.40	5.09	5.38	5.47	1.22
10. old and infirm persons, children & students	98.79	68.73	18.05	4.17	1.02	0.82	1.72	13.62	31.54	59.11	39.98
(iii) sub-total :	99.19	94.24	88.76	85.62	82.91	75.58	73.26	77.84	79.57	89.58	86.95
11. not recorded	0.71	0.96	0.59	0.23	0.41	0.67	0.22	0.15	0.00	0.26	0.56
(iv) total :	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(1,985)	(2,101)	(338)	(863)	(983)	(1,335)	(931)	(668)	(279)	(384)	(9,867)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (12) : PERCENTAGE DISTRIBUTION OF FEMALES BY AGE-GROUPS AND INDUSTRIAL STATUS

(b) base : females of each industrial status

industrial status	age-groups (in years)										total	no. of sample persons
	0–6	7–15	16–17	18–21	22–26	27–36	37–46	47–56	57–61	62 and above		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
gainfully occupied persons:												
1. employers	0.00	0.00	0.00	14.29	14.28	28.57	28.57	0.00	14.29	0.00	100.00	(7)
2. employees	0.32	8.71	2.26	9.84	12.90	26.94	20.64	12.10	4.03	2.26	100.00	(620)
3. own-account workers	0.00	3.54	1.61	6.43	12.54	25.40	25.40	11.90	6.75	6.43	100.00	(311)
4. unpaid household labour	0.00	14.48	6.79	8.15	15.39	23.98	11.76	13.57	4.07	1.81	100.00	(221)
(i) sub-total :	0.17	8.37	2.93	8.63	13.29	25.97	20.28	12.25	4.83	3.28	100.00	(1,159)
unemployed persons seeking employment:												
5. for the first time	0.00	11.77	5.88	41.18	8.82	8.82	17.65	2.94	2.94	0.00	100.00	(34)
6. not for the first time	0.00	0.00	0.00	20.51	17.95	33.33	15.39	10.26	0.00	2.56	100.00	(39)
(ii) sub-total :	0.00	5.48	2.74	30.14	13.70	21.92	16.44	6.85	1.37	1.36	100.00	(73)
persons not in the labour force:												
7. family members engaged in domestic work only	0.18	11.96	5.33	15.53	17.76	21.64	14.39	8.59	2.54	2.08	100.00	(4,482)
8. rentiers and pensioners	0.00	0.00	0.00	3.03	0.00	18.18	24.24	30.31	15.15	9.09	100.00	(33)
9. beggars and persons living on charities & remittances	0.00	0.00	0.00	5.00	7.50	18.33	10.84	28.33	12.50	17.50	100.00	(120)
10. old and infirm persons, children and students	49.71	36.60	1.55	0.91	0.25	0.28	0.41	2.31	2.23	5.75	100.00	(3,945)
(iii) sub-total :	22.95	23.07	3.50	8.61	9.50	11.76	7.95	6.06	2.59	4.01	100.00	(8,580)
11. not recorded	25.45	36.36	3.64	3.64	7.27	16.36	3.64	1.82	0.00	1.82	100.00	(55)
(iv) total :	20.12	21.29	3.43	8.75	9.96	13.53	9.43	6.77	2.83	3.89	100.00	(9,867)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (13) : PERCENTAGE DISTRIBUTION OF POPULATION BY INDUSTRIAL STATUS AND GENERAL EDUCATION

(a) base : persons of each general education standard

industrial status	general education						total
	illiterate	literate but below matric	matric	intermediate	graduate and above	not recorded	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
gainfully occupied persons:							
1. employers	0.13	0.79	1.85	1.95	2.31	2.22	0.47
2. employees	13.24	20.79	45.60	40.26	57.69	4.44	17.59
3. own-account workers	10.47	12.94	8.81	12.34	16.92	6.67	11.40
4. unpaid household labour	3.19	2.16	1.55	3.90	0.00	4.45	2.74
(i) sub-total :	27.03	36.68	57.81	58.45	76.92	17.78	32.20
unemployed persons seeking employment:							
5. for the first time	0.33	1.84	6.49	5.19	4.62	0.00	1.16
6. not for the first time	0.64	2.28	4.17	5.19	3.08	2.22	1.43
(ii) sub-total :	0.97	4.12	10.66	10.38	7.70	2.22	2.59
persons not in the labour force:							
7. family members engaged in domestic work only	25.11	20.00	6.65	3.90	3.85	20.00	22.28
8. rentiers and pensioners	0.26	0.88	2.63	1.30	5.38	0.00	0.61
9. beggars and persons living on charities and remittances	0.90	0.75	0.46	0.00	0.00	2.22	0.82
10. old and infirm persons, children and students	45.06	37.19	21.33	25.32	6.15	55.56	40.95
(iii) sub-total :	71.33	58.82	31.07	30.52	15.38	77.78	64.66
11. not recorded	0.67	0.38	0.46	0.65	0.00	2.22	0.55
(iv) total :	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(11,704)	(7,725)	(647)	(154)	(130)	(45)	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (13) : PERCENTAGE DISTRIBUTION OF POPULATION BY INDUSTRIAL STATUS AND GENERAL EDUCATION

(b) base : persons of each industrial status

industrial status	general education						total	no. of sample persons
	illiterate	literate but below matric	matric	intermediate	graduate and above	not recorded		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
gainfully occupied persons:								
1. employers	15.79	64.21	12.63	3.16	3.16	1.05	100.00	(95)
2. employees	43.17	44.73	8.22	1.73	2.09	0.06	100.00	(3,590)
3. own-account workers	52.66	42.99	2.45	0.82	0.95	0.13	100.00	(2,326)
4. unpaid household labour	66.84	29.93	1.79	1.08	0.00	0.36	100.00	(558)
(i) sub-total :	48.16	43.14	5.69	1.37	1.52	0.12	100.00	(6,569)
unemployed persons seeking employment:								
5. for the first time	16.46	59.91	17.72	3.38	2.53	0.00	100.00	(237)
6. not for the first time	25.77	60.49	9.28	2.75	1.37	0.34	100.00	(291)
(ii) sub-total :	21.59	60.23	13.07	3.03	1.89	0.19	100.00	(528)
persons not in the labour force:								
7. family members engaged in domestic work only	64.63	33.98	0.95	0.13	0.11	0.20	100.00	(4,547)
8. rentiers and pensioners	24.80	54.40	13.60	1.60	5.60	0.00	100.00	(125)
9. beggars and persons living on charities and remittances	62.87	34.73	1.80	0.00	0.00	0.60	100.00	(167)
10. old and infirm persons, children and students	63.10	34.38	1.65	0.47	0.10	0.30	100.00	(8,356)
(iii) sub-total :	63.26	34.44	1.52	0.36	0.15	0.27	100.00	(13,195)
11. not recorded	69.93	25.66	2.65	0.88	0.00	0.88	100.00	(113)
12. total	57.36	37.86	3.17	0.75	0.64	0.22	100.00	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (14) : PERCENTAGE DISTRIBUTION OF POPULATION BY HOUSEHOLD EXPENDITURE CLASSES AND INDUSTRIAL STATUS

(a) base : persons of each household expenditure class

industrial status	expenditure classes (in rupees per month)								total
	upto 50	51–100	101–200	201–300	301–500	501–1000	above 1000	not recorded	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
gainfully occupied persons:									
1. employers	0.22	0.27	0.74	0.81	0.46	1.40	5.66	0.00	0.47
2. employees	20.92	17.30	15.89	13.96	14.98	23.25	27.36	19.14	17.59
3. own-account workers	16.39	12.09	8.02	6.70	7.55	6.51	2.83	12.96	11.40
4. unpaid household labour	2.36	3.20	2.55	3.20	1.86	0.00	6.60	0.00	2.74
(i) sub-total :	39.89	32.86	27.20	24.67	24.85	31.16	42.45	32.10	32.20
unemployed person seeking employment:									
5. for the first time	0.75	1.02	1.44	1.75	2.09	2.33	0.00	0.00	1.16
6. not for the first time	1.59	1.49	1.38	1.25	0.70	0.00	0.00	3.09	1.43
(ii) sub-total :	2.34	2.51	2.82	3.00	2.79	2.33	0.00	3.09	2.59
persons not in the labour force:									
7. family members engaged in domestic work only	20.89	23.04	23.19	21.12	22.42	17.21	15.09	24.07	22.28
8. rentiers and pensioners	0.52	0.41	0.80	1.00	0.70	1.40	0.00	1.85	0.61
9. beggars and persons living on charities and remittances	2.16	0.49	0.35	0.19	0.23	0.00	0.00	0.62	0.82
10. old and infirm persons, children and students	33.68	40.26	44.86	49.44	48.43	47.43	42.46	38.27	40.95
(iii) sub-total :	57.25	64.20	69.20	71.75	71.78	66.04	57.55	64.81	64.66
11. not recorded	0.52	0.43	0.78	0.58	0.58	0.47	0.00	0.00	0.55
12. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
number of sample persons	(4,954)	(7,375)	(5,136)	(1,596)	(861)	(215)	(106)	(162)	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (14) : PERCENTAGE DISTRIBUTION OF POPULATION BY HOUSEHOLD EXPENDITURE CLASSES AND INDUSTRIAL STATUS

(b) base : persons of each industrial status

industrial status	expenditure classes (in rupees per month)								total	no. of sample persons
	up to 50	51–100	101–200	201–300	301–500	501–1000	above 1000	not recorded		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
gainfully occupied persons:										
1. employers	11.58	21.05	40.00	13.68	4.21	3.16	6.32	0.00	100.00	(95)
2. employees	28.86	35.55	22.73	6.21	3.59	1.39	0.81	0.86	100.00	(3,590)
3. own-account workers	34.91	38.36	17.71	4.60	2.79	0.60	0.13	0.90	100.00	(2,326)
4. unpaid household labour	20.97	42.29	23.48	9.14	2.87	0.00	1.25	0.00	100.00	(558)
(i) sub-total :	30.08	36.89	21.27	6.00	3.26	1.02	0.69	0.79	100.00	(6,569)
unemployed persons seeking employment:										
5. for the first time	15.61	31.66	31.22	11.81	7.59	2.11	0.00	0.00	100.00	(237)
6. not for the first time	27.15	37.80	24.40	6.87	2.06	0.00	0.00	1.72	100.00	(291)
(ii) sub-total :	21.97	35.03	27.46	9.09	4.55	0.95	0.00	0.95	100.00	(528)
persons not in the labour force:										
7. family members engaged in domestic work only	22.76	37.38	26.19	7.41	4.24	0.81	0.35	0.86	100.00	(4,547)
8. rentiers and pensioners	20.80	24.00	32.80	12.80	4.80	2.40	0.00	2.40	100.00	(125)
9. beggars and persons living on charities and remittances	64.06	21.56	10.78	1.80	1.20	0.00	0.00	0.60	100.00	(167)
10. old and infirm persons, children and students	19.98	35.54	27.57	9.44	4.99	1.22	0.54	0.74	100.00	(8356)
(iii) sub-total :	21.49	35.88	26.93	8.68	4.68	1.08	0.46	0.80	100.00	(13,195)
11. not recorded	23.01	28.32	35.41	7.96	4.42	0.88	0.00	0.00	100.00	(113)
12. total :	24.28	36.14	25.17	7.82	4.22	1.05	0.52	0.80	100.00	(20,405)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (15) : PERCENTAGE DISTRIBUTION OF PERSONS IN THE LABOUR FORCE BY INDUSTRIAL STATUS AND TYPES OF MIGRANTS AND NON-MIGRANTS

(a) base : persons of each industrial status

type of migrants or non-migrants	industrial status								total
	gainfully occupied persons					unemployed persons seeking employment			
	employers	employees	own-account workers	unpaid house-hold labour	sub-total	for the first time	not for the first time	sub-total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
non-migrants									
1. local persons	69.47	46.46	61.61	70.61	54.21	43.46	47.77	45.84	53.59
2. temporary visitors	0.00	5.16	3.05	2.15	4.08	5.06	4.47	4.73	4.13
(i) sub-total :	69.47	51.62	64.66	72.76	58.29	48.52	52.24	50.57	57.72
migrants from									
3. rural area of the same state	4.21	21.67	15.35	10.93	18.27	10.97	12.37	11.74	17.78
4. urban area of the same state	5.26	7.91	3.78	5.20	6.18	2.95	4.12	3.60	5.99
5. rural area of a different state	3.16	7.02	2.92	1.97	5.08	2.11	2.40	2.27	4.88
6. urban area of a different state	0.00	1.42	0.35	0.54	0.94	2.11	1.03	1.52	0.99
7. East Pakistan as displaced persons	8.42	5.24	5.07	2.15	4.96	12.66	9.62	10.98	5.41
8. West Pakistan as displaced persons	3.16	2.98	5.76	1.43	3.84	4.22	6.19	5.30	3.94
9. other places	4.21	0.89	0.86	0.54	0.90	0.42	0.69	0.57	0.87
(ii) sub-total :	28.42	47.13	34.09	22.76	40.17	35.44	36.42	35.98	39.86
10. not recorded	2.11	1.25	1.25	4.48	1.54	16.04	11.34	13.45	2.42
11. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(95)	(3,590)	(2,326)	(558)	(6,569)	(237)	(291)	(528)	(7,097)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (15) : PERCENTAGE DISTRIBUTION OF PERSONS IN THE LABOUR FORCE BY INDUSTRIAL STATUS AND TYPES OF MIGRANTS AND NON-MIGRANTS

(b) base : persons in the labour force belonging to each type of migrants or non-migrants

type of migrants or non-migrnta	industrial status								total	no. of sample persons
	gainfully occupied persons					unemployed persons seeking employment				
	employers	employees	own-account workers	unpaid house-hold labour	sub-total	for the first time	not for the first time	sub-total		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
non-migrants:										
1. local persons	1.74	43.86	37.68	10.36	93.64	2.71	3.65	6.36	100.00	(3,803)
2. temporary visitors	0.00	63.14	24.23	4.10	91.47	4.09	4.44	8.53	100.00	(293)
(i) sub-total :	1.61	45.24	36.72	9.91	93.48	2.81	3.71	6.52	100.00	(4,096)
migrants from:										
3. rural area of the same state	0.32	61.65	28.29	4.83	95.09	2.06	2.85	4.91	100.00	(1,262)
4. urban area of the same state	1.18	66.82	20.71	6.82	95.53	1.65	2.82	4.47	100.00	(425)
5. rural area of a different state	0.87	72.83	19.65	3.18	96.53	1.45	2.02	3.47	100.00	(346)
6. urban area of a different state	0.00	72.86	11.43	4.28	88.57	7.14	4.29	11.43	100.00	(70)
7. East Pakistan as displaced persons	2.08	48.96	30.73	3.13	84.90	7.81	7.29	15.10	100.00	(384)
8. West Pakistan as displaced persons	1.07	38.21	47.86	2.86	90.00	3.57	6.43	10.00	100.00	(280)
9. other places	6.45	51.61	32.26	4.84	95.16	1.61	3.23	4.84	100.00	(62)
(ii) sub-total :	0.95	59.81	28.03	4.49	93.28	2.97	3.75	6.72	100.00	(2,829)
10. not recorded	1.16	26.16	16.86	14.54	58.72	22.09	19.19	41.28	100.00	(172)
11. total :	1.34	50.59	32.77	7.86	92.56	3.34	4.10	7.44	100.00	(7,097)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (16) : PERCENTAGE DISTRIBUTION OF MIGRANTS IN THE LABOUR FORCE BY INDUSTRIAL STATUS AND AREA FROM WHERE MIGRATED

base : migrants of each industrial status

area from where migrated	industrial status								total
	gainfully occupied persons					unemployed persons seeking employment			
	employers	employees	own-account workers	unpaid house-hold labour	sub-total	for the first time	not for the first time	sub-total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. rural area of the same state	14.81	45.99	45.02	48.04	45.47	30.96	33.96	32.62	44.62
2. urban area of the same state	18.52	16.79	11.10	22.83	15.38	8.33	11.32	10.00	15.02
3. rural area of a different state	11.11	14.89	8.58	8.66	12.66	5.95	6.60	6.32	12.23
4. urban area of a different state	0.00	3.01	1.00	2.36	2.35	5.95	2.83	4.21	2.47
5. East Pakistan as displaced persons	29.64	11.11	14.88	9.45	12.35	35.72	26.42	30.53	13.57
6. West Pakistan as displaced persons	11.11	6.32	16.90	6.30	9.55	11.90	16.98	14.74	9.90
7. other places	14.81	1.89	2.52	2.36	2.24	1.19	1.89	1.58	2.19
8. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(27)	(1,629)	(793)	(127)	(2,639)	(84)	(106)	(190)	(2,829)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (17) : PERCENTAGE DISTRIBUTION OF MIGRANTS IN THE LABOUR FORCE BY AREA FROM WHERE MIGRATED AND REASONS FOR MIGRATION

(a) base : migrants in the labour force having each type of reason for migration

area from where migrated	reasons of migration					total
	unemployment	partial unemployment	uncongenial occupation	further studies	other reasons	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. rural area of the same state	59.76	53.85	56.84	54.76	29.59	44.62
2. urban area of the same state	13.69	15.38	20.00	11.91	15.88	15.02
3. rural area of a different state	18.68	24.62	7.37	2.38	6.16	12.23
4. urban area of a different state	2.03	0.00	2.11	7.14	2.97	2.47
5. East Pakistan as displaced persons	3.64	1.54	11.58	19.05	23.20	13.57
6. West Pakistan as displaced persons	0.17	0.77	1.05	0.00	20.02	9.90
7. other places	2.03	3.84	1.05	4.76	2.18	2.19
8. total	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(1,183)	(130)	(95)	(42)	(1,379)	(2,829)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (17) : PERCENTAGE DISTRIBUTION OF MIGRANTS IN THE LABOUR FORCE BY AREA FROM WHERE MIGRATED AND REASONS FOR MIGRATION

(b) base : migrants in the labour force from each area of migration

area from where migrated	reasons of migration						no. of sample persons
	unemployment	partial unemployment	uncongenial occupation	further studies	other reasons	total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. rural area of the same state	56.02	5.55	4.28	1.82	32.33	100.00	(1,262)
2. urban area of the same state	38.12	4.70	4.47	1.18	15.53	100.00	(425)
3. rural area of a different state	63.87	9.25	2.02	0.29	24.57	100.00	(346)
4. urban area of a different state	34.29	0.00	2.86	4.28	58.57	100.00	(70)
5. East Pakistan as displaced persons	11.20	0.52	2.86	2.08	83.34	100.00	(384)
6. West Pakistan as displaced persons	0.71	0.36	0.36	0.00	98.57	100.00	(280)
7. other places	38.71	8.06	1.61	3.22	48.40	100.00	(62)
8. total	41.82	4.60	3.36	1.48	48.74	100.00	(2,829)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (18) : PERCENTAGE DISTRIBUTION OF MIGRANTS IN THE LABOUR FORCE BY AREA FROM WHERE MIGRATED AND RANGES OF DURATION OF MIGRATION

(a) base : migrants in the labour force having each range of duration of migration

area from where migrated	ranges of duration of migration								total
	up to 6 months	6 months– 1 year	1 year– 2 years	2 years– 4 years	4 years– 6 years	6 years– 10 years	10 years– 15 years	above 15 years	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. rural area of the same state	43.33	40.00	41.55	30.94	20.79	55.05	59.91	60.96	44.62
2. urban area of the same state	25.83	39.09	19.32	9.46	9.79	15.38	19.40	14.01	15.02
3. rural area of a different state	10.00	11.82	16.43	9.21	6.01	12.74	11.64	17.64	12.23
4. urban area of a different state	6.67	0.91	3.38	1.28	1.72	1.92	3.45	2.98	2.47
5. East Pakistan as displaced persons	11.67	6.36	18.84	44.25	22.51	4.81	0.00	0.00	13.57
6. West Pakistan as displaced persons	0.83	0.00	0.48	4.60	37.80	9.62	0.00	0.00	9.90
7. other places	1.67	1.82	0.00	0.26	1.38	0.48	5.60	4.41	2.19
8. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(120)	(110)	(207)	(391)	(582)	(416)	(232)	(771)	(2,829)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (19) : PERCENTAGE DISTRIBUTION OF MIGRANTS IN THE LABOUR FORCE BY REASONS FOR MIGRATION AND RANGES OF DURATION OF MIGRATION

(b) base : migrants in the labour force having each type of reason for migration

reason for migration	ranges of duration of migration								total	no. of sample persons
	up to 6 months	6 months–1 year	1 year–2 years	2 years–4 years	4 years–6 years	6 years–10 years	10 years–15 years	above 15 years		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1. unemployment	4.99	4.99	8.03	10.65	10.57	19.10	8.79	32.88	100.00	(1,183)
2. partial unemployment	2.31	6.15	6.92	4.62	12.31	17.69	10.77	39.23	100.00	(130)
3. uncongenial occupation	6.32	5.26	8.42	10.54	12.63	16.83	12.63	27.37	100.00	(95)
4. further studies	4.76	0.00	4.76	19.05	19.05	14.29	7.14	30.95	100.00	(42)
5. other reasons	3.63	2.76	6.74	17.48	30.53	10.51	7.18	21.17	100.00	,379)
6. total	4.24	3.89	7.32	13.82	20.57	14.70	8.20	27.26	100.00	(2,829)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (20) : PERCENTAGE DISTRIBUTION OF GAINFULLY OCCUPIED PERSONS BY INDUSTRY AND INDUSTRIAL STATUS

industry	(a) base : gainfully occupied persons belonging to each industry						(b) base : persons of each industrial status				
	industrial status						industrial status				
	employers	employees	own-account workers	unpaid household labour	total	no. of sample persons	employers	employees	own-account workers	unpaid household labour	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(2)	(3)	(4)	(5)	(6)
1. agriculture, livestock, forestry, fishing and hunting	2.96	24.53	46.63	25.88	100.00	(742)	23.16	5.07	14.88	34.40	11.30
2. mining and quarrying	0.00	96.64	3.36	0.00	100.00	(119)	0.00	3.20	0.17	0.00	1.81
3. manufacture	1.77	60.07	30.04	8.12	100.00	(2,094)	38.95	35.05	27.04	30.46	31.87
4. construction	0.51	56.87	40.10	1.52	100.00	(197)	1.05	3.18	3.40	0.54	3.00
5. electricity, water, gas and sanitary services	0.00	80.68	14.77	4.55	100.00	(88)	0.00	1.98	0.56	0.72	1.34
6. distributive services	2.29	25.32	62.23	10.16	100.00	(1,181)	28.42	8.33	31.60	21.51	17.98
7. transport, storage and communication	0.00	71.07	28.48	0.45	100.00	(660)	0.00	13.06	8.08	0.54	10.05
9. other services not covered above	0.48	73.48	22.21	3.83	100.00	(1,463)	7.37	29.94	13.97	10.04	22.27
9. not recorded	4.00	28.00	28.00	40.00	100.00	(25)	1.05	0.19	0.30	1.79	0.38
10. total	1.45	54.65	25.41	8.49	100.00	(6,569)	100.00	100.00	100.00	100.00	100.00

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (21) : PERCENTAGE DISTRIBUTION OF GAINFULLY OCCUPIED PERSONS BY INDUSTRY AND INCOME GROUPS

(a) base : gainfully occupied persons of each income group

industry	income groups (in rupees per month)										total
	up to 25	26–50	51–100	101–150	151–200	201–300	301–500	501–1000	above 1000	not recorded	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1. agriculture, livestock, forestry, fishing and hunting	20.38	6.20	2.82	1.63	4.93	5.43	4.55	5.88	5.00	32.67	11.30
2. mining and quarrying	0.51	5.51	0.96	0.81	0.00	0.00	0.00	0.00	0.00	0.24	1.81
3. manufacture	32.21	29.47	39.62	28.72	24.65	13.95	20.45	17.65	20.00	24.26	31.87
4. construction	4.20	3.50	1.92	1.63	1.41	1.55	2.27	5.88	0.00	1.49	3.00
5. electricity, water, gas, and sanitary services	1.11	2.00	1.03	1.83	0.70	1.55	0.00	0.00	0.00	0.99	1.34
6. distributive services	14.00	20.40	16.92	19.35	25.35	25.58	45.46	29.41	70.00	21.04	17.98
7. transport, storage and communication	5.64	8.64	15.77	21.39	15.49	13.18	6.82	0.00	0.00	1.73	10.05
8. other services not covered above	21.77	23.97	20.83	24.64	27.47	38.76	20.45	41.18	5.00	14.11	22.27
9. not recorded	0.18	0.31	0.13	0.00	0.00	0.00	0.00	0.00	0.00	3.47	0.38
10. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(2,164)	(1,598)	(1,560)	(491)	(142)	(129)	(44)	(17)	(20)	(404)	(6,569)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (21) : PERCENTAGE DISTRIBUTION OF GAINFULLY OCCUPIED PERSONS BY INDUSTRY AND INCOME GROUPS

(b) base : gainfully occupied persons belonging to each industry

industry	income groups (in rupees per month)										total	no. of sample persons
	up to 25	26–50	51–100	101–150	151–200	201–300	301–500	501–1000	above 1000	not recorded		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. agriculture, livestock, forestry, fishing and hunting	59.44	13.34	5.93	1.08	0.94	0.94	0.27	0.13	0.13	17.80	100.00	(742)
2. mining and quarrying	9.24	73.95	12.61	3.36	0.00	0.00	0.00	0.00	0.00	0.84	100.00	(119)
3. manufacture	33.29	22.49	29.52	6.73	1.67	0.86	0.43	0.14	0.19	4.68	100.00	(2,094)
4. construction	46.18	28.42	15.23	4.06	1.02	1.02	0.51	0.51	0.00	3.05	100.00	(197)
5. electricity, water, gas and sanitary services	27.27	36.36	18.18	10.23	1.14	2.27	0.00	0.00	0.00	4.55	100.00	(88)
6. distributive services	25.67	27.60	22.36	8.04	3.05	2.79	1.69	0.42	1.19	7.19	100.00	(1,181)
7. transport, storage and communication	18.48	20.92	37.27	15.91	3.33	2.58	0.45	0.00	0.00	1.06	100.00	(660)
8. other service not covered above	32.19	26.18	22.21	8.27	2.67	3.42	0.62	0.48	0.07	3.89	100.00	(1,463)
9. not recorded	16.00	20.00	8.00	0.00	0.00	0.00	0.00	0.00	0.00	56.00	100.00	(25)
10. total	32.94	24.33	23.75	7.47	2.16	1.96	0.68	0.26	0.30	6.15	100.00	(6,569)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (22) : PERCENTAGE DISTRIBUTION OF EMPLOYEES BY GENERAL EDUCATION AND TYPES OF EMPLOYER

(a) base : employees having each type of employer

general education	types of employer			total
	government	non-government	not recorded	
(1)	(2)	(3)	(4)	(5)
1. illiterate	20.53	49.57	63.29	43.17
2. literate but below matric	48.91	43.76	34.18	44.73
3. matric	19.81	4.81	2.53	8.22
4. intermediate	4.23	1.01	0.00	1.73
5. graduate and above	6.52	0.78	0.00	2.09
6. not recorded	0.00	0.07	0.00	0.06
7. total	100.00	100.00	100.00	100.00
no. of sample persons	(828)	(2,683)	(79)	(3,590)

(b) base : employees of each general education standard

general education	types of employer			total	no. of sample persons
	government	non-government	not recorded		
(1)	(6)	(7)	(8)	(9)	(10)
1. illiterate	10.97	85.81	3.22	100.00	(1,550)
2. literate but below matric	25.22	73.10	1.68	10000	(1,606)
3. matric	55.59	43.73	0.68	100.00	(295)
4. intermediate	56.45	43.55	0.00	100.00	(62)
5. graduate and above	72.00	28.00	0.00	100.00	(75)
6. not recorded	0.00	100.00	0.00	100.00	(2)
7. total	23.06	74.74	2.20	100.00	(3,590)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (23) : PERCENTAGE DISTRIBUTION OF GAINFULLY OCCUPIED PERSONS IN EACH INDUSTRY BY INTENSITY OF EMPLOYMENT

industry	intensity of employment								total
	nil	less than quarter	quarter	half	three quarters	sub-total (cols. 2–6)	full	not recorded	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. agriculture, livestock, forestry, fishing and hunting	4.26	4.68	3.40	11.49	4.26	28.09	71.48	0.43	100.00
2. mining and quarrying	0.84	0.84	0.00	7.56	2.52	11.76	88.24	0.00	100.00
3. manufacture	2.38	3.05	4.76	10.86	12.24	33.29	65.90	0.81	100.00
4. construction	6.57	14.65	16.67	16.67	10.61	65.17	34.33	0.50	100.00
5. electricity, water, gas and sanitary services	2.22	1.11	2.22	4.44	2.22	12.21	87.79	0.00	100.00
6. distributive services	1.78	2.03	2.11	6.18	6.85	18.95	80.71	0.34	100.00
7. transport, storage and communication	2.27	1.21	5.00	10.91	8.18	27.57	72.13	0.30	100.00
8. other services not covered above	3.02	3.50	1.99	5.15	3.91	17.57	82.29	0.14	100.00
9. total	2.86	3.23	3.74	8.76	7.73	26.32	72.93	0.75	100.00

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (24) : PERCENTAGE DISTRIBUTION OF GAINFULLY OCCUPIED PERSONS BY INTENSITY OF EMPLOYMENT AND INCOME

income group (in rupees)	intensity of employment								total
	nil	less than quarter	quarter	half	three quarters	sub-total (cols. 2–6)	full	not recorded	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. up to 25	62.76	85.85	73.58	57.30	31.50	56.13	24.93	—	32.94
2. 26— 50	5.85	5.66	15.85	27.08	30.51	21.56	25.60	—	24.33
3. 51— 100	4.26	1.89	5.28	8.51	24.61	11.50	28.44	—	23.75
4. 101— 150	2.13	0.00	0.41	1.04	7.87	2.95	9.18	—	7.47
5. 151— 200	1.06	0.47	0.81	0.00	1.38	0.69	2.69	—	2.16
6. 201— 300	1.60	0.47	0.41	0.35	0.39	0.52	2.50	—	1.96
7. 301— 500	0.00	0.00	0.00	0.17	0.00	0.06	0.92	—	0.68
8. 501—1000	0.00	0.00	0.00	0.17	0.00	0.06	0.31	—	0.26
9. above 1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	—	0.30
10. not recorded	22.34	5.66	3.66	5.38	3.74	6.53	5.01	100.00	6.15
11. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample persons	(188)	(212)	(246)	(576)	(508)	(1,730)	(4,793)	(49)	(6,572)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (25) : PERCENTAGE DISTRIBUTION OF UNDEREMPLOYED GAINFULLY OCCUPIED PERSONS BY INTENSITY OF EMPLOYMENT AND REASON FOR UNDEREMPLOYMENT

reason for underemployment	intensity of employment					total	no. of sample persons
	nil	less than quarter	quarter	half	three quarters		
(1)	(2)	(3)	(4)	(6)	(6)	(7)	(8)
(a) base : underemployed gainfully occupied persons having each type of reason for underemployment							
1. disability to accept work for illness	24.68	9.49	10.13	22.78	32.92	100.00	(316)
2. domestic reasons	6.93	7.92	11.88	22.77	50.50	100.00	(101)
3. fall in production due to lack of materials and equipments	6.78	5.08	11.86	27.12	49.16	100.00	(59)
4. fall in demand	1.97	12.47	18.16	39.83	27.57	100.00	(457)
5. industrial disputes	0.00	0.00	0.00	0.00	100.00	100.00	(2)
6. slack season	11.02	19.77	17.51	33.90	17.80	100.00	(354)
7. other reasons	10.55	10.09	11.47	37.39	30.50	100.00	(436)
8. not recorded	100.00	0.00	0.00	0.00	0.00	100.00	(5)
9. total	10.86	12.24	14.21	33.32	29.37	100.00	(1,730)
(b) base : underemployed gainfully occupied persons with each degree of intensity of employment							
1. disability to accept work for illness	41.49	14.15	13.01	12.50	20.47	18.26	
2. domestic reasons	3.72	3.77	4.88	3.99	10.04	5.83	
3. fall in production due to lack of materials and equipments	2.13	1.42	2.85	2.78	5.71	3.40	
4. fall in demand	4.79	26.89	33.73	31.60	24.81	26.43	
5. slack season	20.74	33.02	25.20	20.83	12.40	20.46	
6. industrial disputes	0.00	0.00	0.00	0.00	0.39	0.12	
7. other reasons	24.47	20.75	20.33	28.30	26.18	25.21	
8. not recorded	2.66	0.00	0.00	0.00	0.00	0.29	
9. total	100.00	100.00	100.00	100.00	100.00	100.00	
no. of sample persons	(188)	(212)	(246)	(576)	(508)	(1,730)	

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (26) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY TYPES AND EDUCATION

general education	technical or professional qualification[1]	unemployed persons seeking employment		
		for the first time	not for the first time	total
(1)	(2)	(3)	(4)	(5)
1. illiterate	A	12.66	16.49	14.77
	B	3.80	9.28	6.82
	i) sub-total	16.46	25.77	21.59
2. literate but below matric	A	43.45	28.87	35.42
	B	16.46	29.21	23.48
	C	0.00	2.41	1.33
	ii) sub-total	59.91	60.49	60.23
3. matric	A	13.51	3.10	7.76
	B	2.95	3.78	3.41
	C	0.84	1.37	1.14
	D	0.42	1.03	0.76
	iii) sub-total	17.72	9.28	13.07
4. intermediate	A	2.12	2.07	2.08
	B	0.42	0.34	0.38
	C	0.84	0.00	0.38
	D	0.00	0.34	0.19
	iv) sub-total	3.38	2.75	3.03
5. graduate and above	A	1.69	1.03	1.32
	B	0.42	0.00	0.19
	C	0.00	0.00	0.00
	D	0.42	0.34	0.38
	v) sub-total	2.53	1.37	1.89
6. not recorded		0.00	0.34	0.19
7. total	A	73.43	51.56	61.35
	B	24.05	42.61	34.28
	C	1.68	3.78	2.85
	D	0.84	1.71	1.33
	not recorded	0.00	0.34	0.19
	vi) total	100.00	100.00	100.00
no. of sample persons		(237)	(291)	(528)

number of sample towns = 23 ; number of sample blocks = 212 ; number of sample households = 4,446

[1] *Technical or professional qualification:*

A. no technical or professional qualification;

B. professional or technical skill without any degree or equivalent diploma and not covered by C:

C. certificate or diploma in typing, stenography, nursing, midwifery or pharmacy:

D. degree or equivalent diploma in any technology or profession.

TABLE (27) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY TYPES AND INCOME EXPECTED

range of income expected in rupees per month	unemployed persons seeking employment		
	for the first time	not for the first time	total
(1)	(2)	(3)	(4)
1. up to 25	8.02	8.93	8.52
2. 26— 50	29.54	29.55	29.55
3. 51—100	42.60	45.03	43.94
4. 101—150	14.77	10.65	12.50
5. 151—200	1.69	1.72	1.70
6. above 200	0.00	0.34	0.19
7. not recorded	3.38	3.78	3.60
8. total	100.00	100.00	100.00

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (28) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY GENERAL EDUCATION AND INCOME EXPECTED

general education	ranges of income expected (in rupees)							total	no. of sample persons
	up to 25	26–50	51–100	101–150	151–200	above 200	not recorded		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(a) base : unemployed persons with each range of income expected									
1. illiterate	57.78	38.00	11.44	1.47	0.00	0.00	15.79	21.59	
2. literate but below matric	42.22	58.00	69.07	47.06	44.45	0.00	68.42	60.23	
3. matric	0.00	3.33	17.37	27.94	22.22	0.00	10.53	13.07	
4. intermediate	0.00	0.00	1.27	16.18	11.11	0.00	5.26	3.03	
5. graduate and above	0.00	0.00	0.85	7.35	22.22	100.00	0.00	1.89	
6. not recorded	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.19	
7. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	
no. of sample persons	(45)	(15)	(236)	(68)	(9)	(1)	(19)	(528)	
(b) base : unemployed persons of each general education standard									
1. illiterate	22.81	50.00	23.68	0.88	0.00	0.00	2.63	100.00	(114)
2. literate but below matric	5.97	27.36	51.26	10.06	1.26	0.00	4.09	100.00	(318)
3. matric	0.00	7.25	59.41	27.54	2.90	0.00	2.90	100.00	(69)
4. intermediate	0.00	0.00	18.75	68.75	6.25	0.00	6.25	100.00	(16)
5. graduate and above	0.00	0.00	20.00	50.00	20.00	10.00	0.00	100.00	(10)
6. not recorded	0.00	100.00	0.00	0.00	0.00	0.00	0.00	100.00	(1)
7. total	8.52	29.55	43.94	12.50	1.70	0.19	3.60	100.00	(528)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (29) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY NATURE OF EMPLOYMENT SOUGHT AND GENERAL EDUCATION

(a) base : unemployed persons with each type of employment sought

nature of employment sought	general education							no. of sample persons
	illiterate	literate but below matric	matric	inter-mediate	graduate and above	not recorded	total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
as own-account worker:								
1. agricultural production	50.00	50.00	0.00	0.00	0.00	0.00	100.00	(4)
2. animal husbandry	25.00	50.00	25.00	0.00	0.00	0.00	100.00	(4)
3. manufacture and handicraft	39.47	57.90	2.63	0.00	0.00	0.00	100.00	(38)
4. transport services	0.00	91.67	8.33	0.00	0.00	0.00	100.00	(12)
5. trade	13.33	77.78	6.67	0.00	2.22	0.00	100.00	(45)
9. professions and financial transactions	28.00	48.00	12.00	4.00	8.00	0.00	100.00	(25)
(i) sub-total (rows 1–6)	24.22	65.53	7.03	0.78	2.34	0.00	100.00	(128)
as employee:								
7. manual work—unskilled	38.89	59.52	0.79	0.00	0.00	0.80	100.00	(126)
8. ,, ,, —skilled	13.51	78.38	8.11	0.00	0.00	0.00	100.00	(74)
(ii) sub-total (rows 7 and 8)	29.50	66.50	3.50	0.00	0.00	0.50	100.00	(200)
9. non-manual work—administrative, executive and clerical	1.19	35.71	46.43	13.10	3.57	0.00	100.00	(84)
10. ,, ,, —technical and technological	8.70	78.26	4.34	0.00	8.70	0.00	100.00	(23)
11. ,, ,, —medical, surgical etc.	0.00	85.71	14.29	0.00	0.00	0.00	100.00	(7)
12. ,, ,, —legal, teaching, writing and the like	0.00	40.00	35.00	15.00	10.00	0.00	100.00	(20)
13. ,, ,, —fine arts	0.00	100.00	0.00	0.00	0.00	0.00	100.00	(3)
14. ,, ,, —others	35.90	56.41	7.69	0.00	0.00	0.00	100.00	(39)
(iii) sub-total (rows 9–14)	9.66	49.43	28.98	7.95	3.98	0.00	100.00	(176)
(iv) sub-total (rows 7–14)	20.21	58.51	15.43	3.72	1.86	0.27	100.00	(376)
15. not recorded	29.17	58.33	8.33	4.17	0.00	0.00	100.00	(24)
16. total	21.59	60.23	13.07	3.03	1.89	0.19	100.00	(528)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (29) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY NATURE OF EMPLOYMENT SOUGHT AND GENERAL EDUCATION

(b) base : unemployed persons of each general education standard

nature of employment sought	general education						total
	illiterate	literate but below matric	matric	inter-mediate	graduate and above	not recorded	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
as own-account worker:							
1. agricultural production	1.75	0.63	0.00	0.00	0.00	0.00	0.76
2. animal husbandry	0.89	0.63	1.45	0.00	0.00	0.00	0.76
3. manufacture and handicraft	13.16	6.92	1.45	0.00	0.00	0.00	7.20
4. transport services	0.00	3.46	1.45	0.00	0.00	0.00	2.27
5. trade	5.26	11.01	4.35	0.00	10.00	0.00	8.52
6. professions and financial transactions	6.14	3.77	4.35	6.25	20.00	0.00	4.73
(i) sub-total (rows 1–6)	27.20	26.42	13.05	6.25	30.00	0.00	24.24
as employee:							
7. manual work—unskilled	42.98	23.58	1.45	0.00	0.00	100.00	23.85
8. ,, ,, —skilled	8.77	18.24	8.70	0.00	0.00	0.00	14.02
(ii) sub-total (rows 7 and 8)	51.75	41.82	10.15	0.00	0.00	100.00	37.87
9. non-manual work—administrative, executive and clerical	0.88	9.43	56.52	68.75	30.00	0.00	15.90
10. ,, ,, —technical and technological	1.75	5.66	1.45	0.00	20.00	0.00	4.36
11. ,, ,, —medical and surgical	0.00	1.89	1.45	0.00	0.00	0.00	1.33
12. ,, ,, —legal, teaching, writing and the like	0.00	2.52	10.14	18.75	20.00	0.00	3.79
13. ,, ,, —fine arts	0.00	0.94	0.00	0.00	0.00	0.00	0.57
14. ,, ,, —others	12.28	6.92	4.34	0.00	0.00	0.00	7.39
(iii) sub-total (rows 9–14)	14.91	27.36	73.90	87.50	70.00	0.00	33.34
(iv) sub-total (rows 7–14)	66.66	69.18	84.05	87.50	70.00	0.00	71.21
15. not recorded	6.14	4.40	2.90	6.25	0.00	0.00	4.55
16. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00
number of sample persons	(114)	(318)	(69)	(16)	(10)	(1)	(528)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (30) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY TYPES AND NATURE OF EMPLOYMENT SOUGHT

nature of employment sought	unemployed persons seeking employment		
	for the first time	not for the first time	total
(1)	(2)	(3)	(4)
as own-account worker:			
1. agricultural production	0.00	1.37	0.76
2. animal husbandry	0.84	0.69	0.76
3. manufacture and handicrafts	5.49	8.59	7.20
4. transport service	2.11	2.41	2.27
5. trade	5.49	11.00	8.52
6. professions and financial transactions	2.53	6.53	4.73
(i) sub-total (rows 1–6)	16.46	30.59	24.24
as employee:			
7. manual work—unskilled	24.04	23.72	23.85
8. ,, —skilled	12.66	15.12	14.02
(ii) sub-total (rows 7–8)	36.70	38.84	37.87
9. non-manual work—administrative, executive or clerical	21.09	11.68	15.90
10. ,, —technical and technological	3.80	4.81	4.36
11. ,, —medical, surgical etc.	1.27	1.37	1.33
12. ,, —legal, teaching, writing and the like	4.22	3.44	3.79
13. ,, —fine arts	1.27	0.00	0.57
14. ,, —others	11.39	4.12	7.39
(iii) sub-total (rows 9–14)	43.04	25.42	33.34
(iv) sub-total (rows 7–14)	79.74	64.26	71.21
15. not recorded	3.80	5.15	4.55
16. total	100.00	100.00	100.00
no. of sample persons	(237)	(291)	(528)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (31) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY NATURE OF EMPLOYMENT SOUGHT AND INCOME EXPECTED

(a) base ; unemployed persons with each type of employment sought

nature of employment sought	ranges of income expected (in Rs.)							total	no. of sample persons
	up to 25	26–50	51–100	101–150	151–200	above 200	not recorded		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
as own-account worker:									
1. agricultural production	0.00	25.00	75.00	0.00	0.00	0.00	0.00	100.00	(4)
2. animal husbandry	0.00	50.00	25.00	0.00	25.00	0.00	0.00	100.00	(4)
3. manufacture and handicraft	18.42	21.05	55.27	5.26	0.00	0.00	0.00	100.00	(38)
4. transport service	0.00	16.67	41.67	41.66	0.00	0.00	0.00	100.00	(12)
5. trade	2.22	26.67	37.78	28.89	4.44	0.00	0.00	100.00	(45)
6. professions and financial transactions	12.00	32.00	28.00	20.00	4.00	4.00	0.00	100.00	(25)
(i) sub-total (rows 1–6)	8.59	25.78	42.19	19.53	3.13	0.78	0.00	100.00	(128)
as employee									
7. manual work—unskilled	15.87	57.94	25.40	0.00	0.00	0.00	0.79	100.00	(126)
8. ,, ,, —skilled	4.06	21.62	68.92	4.05	1.35	0.00	0.00	100.00	(74)
(ii) sub-total (rows 7 and 8)	11.50	44.50	41.50	1.50	0.50	0.00	0.50	100.00	(200)
9. non-manual work—administrative, executive and clerical	2.38	7.14	53.57	32.15	4.76	0.00	0.00	100.00	(84)
10. ,, ,, —technical, and technological	0.00	13.04	73.91	13.05	0.00	0.00	0.00	100.00	(23)
11. ,, ,, —medical, surgical etc.	0.00	42.86	42.86	14.28	0.00	0.00	0.00	100.00	(7)
12. ,, ,, —legal, teaching, writing and the like	5.00	20.00	60.00	15.00	0.00	0.00	0.00	100.00	(20)
13. ,, ,, —fine arts	0.00	33.33	33.33	33.34	0.00	0.00	0.00	100.00	(3)
14. ,, ,, —others	10.26	43.59	38.46	7.69	0.00	0.00	0.00	100.00	(39)
(iii) sub-total (rows 9–14)	3.98	19.32	52.84	21.59	2.27	0.00	0.00	100.00	(176)
(iv) sub-total (rows 7–14)	7.98	32.71	46.81	10.90	1.33	0.00	0.27	100.00	(376)
15. not recorded	16.67	0.00	8.33	0.00	0.00	0.00	75.00	100.00	(24)
16. total	8.52	29.55	43.94	12.50	1.70	0.19	3.60	100.00	(528)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (31) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY NATURE OF EMPLOYMENT SOUGHT AND INCOME EXPECTED

(b) base : unemployed persons with each range of income expected

nature of employment sought	ranges of incomes expected (in Rs.)							
	upto 25	26–50	51–100	101–150	151–200	above 200	not recorded	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
as own-account worker:								
1. agricultural production	0.00	0.64	1.29	0.00	0.00	0.00	0.00	0.76
2. animal husbandry	0.00	1.28	0.43	0.00	11.11	0.00	0.00	0.76
3. manufacture and handicraft	15.56	5.13	9.05	3.03	0.00	0.00	0.00	7.20
4. transport service	0.00	1.28	2.16	7.57	0.00	0.00	0.00	2.27
5. trade	2.22	7.69	7.33	19.70	22.22	0.00	0.00	8.52
6. professions and financial transactions	6.67	5.13	3.01	7.58	11.11	100.00	0.00	4.73
(i) sub-total (rows 1–6)	24.45	21.15	23.27	37.88	44.44	100.00	0.00	24.24
as employee:								
7. manual work—unskilled	44.44	46.79	13.79	0.00	0.00	0.00	5.26	23.85
8. ,, ,, —skilled	6.67	10.26	21.98	4.54	11.11	0.00	0.00	14.02
(ii) sub-total (rows 7 and 8)	51.11	57.05	35.77	4.54	11.11	0.00	5.26	37.87
9. non-manual work—administrative, executive and clerical	4.44	3.85	19.40	40.91	44.45	0.00	0.00	15.90
10. ,, ,, —technical and technological	0.00	1.92	7.34	4.55	0.00	0.00	0.00	4.36
11. ,, ,, —medical, surgical etc.	0.00	1.93	1.29	1.51	0.00	0.00	0.00	1.33
12. ,, ,, —legal, teaching, writing and the like	2.22	2.56	5.17	4.55	0.00	0.00	0.00	3.79
13. ,, ,, —fine arts	0.00	0.64	0.43	1.52	0.00	0.00	0.00	0.57
14. ,, ,, —others	8.89	10.90	6.47	4.54	0.00	0.00	0.00	7.39
(iii) sub-total (rows 9–14)	15.55	21.80	40.10	57.58	44.45	0.00	0.00	33.34
(iv) sub-total (rows 7–14)	66.66	78.85	75.87	62.12	55.56	0.00	5.26	71.21
15. not recorded	8.89	0.00	0.86				94.74	4.55
16. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
number of sample persons	(45)	(156)	(232)	(66)	(9)	(1)	(19)	(528)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

ΓABLE (32) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY TYPES AND ATTITUDE TO REGISTRATION

attitude to registration	unemployed persons seeking employment		
	for the first time	not for the first time	total
(1)	(2)	(3)	(4)
1. registered	24.05	25.09	24.62
not registered due to			
2. ignorance	33.75	35.05	34.47
3. remoteness	1.27	0.69	0.95
4. lack of faith	10.55	10.31	10.42
5. indifference	4.22	4.12	4.17
6. ineligibility	2.55	2.06	2.27
7. other reasons	13.50	14.78	14.20
(i) sub-total (rows 2-7)	65.84	67.01	66.48
8. not recorded	10.11	7.90	8.90
9. total	100.00	100.00	100.00
no. of sample persons	(237)	(291)	(528)

ııumber of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

TABLE (33) : PERCENTAGE DISTRIBUTION OF UNEMPLOYED PERSONS BY GENERAL EDUCATION AND ATTITUDE TO REGISTRATION

attitude to registration	general education							
	illiterate	literate but below matric	matric	inter-mediate	graduate and above	not recorded	total	no. of sample persons
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(a) base : unemployed persons of each general education standard								
1. registered	14.04	24.53	39.13	43.75	20.00	0.00	24.62	
not registered due to								
2. ignorance	56.14	34.90	5.80	12.50	10.00	0.00	34.47	
3. remoteness	0.00	0.63	2.90	6.25	0.00	0.00	0.95	
4. lack of faith	5.26	10.38	15.94	6.25	40.00	0.00	10.42	
5. indifference	3.51	3.77	7.25	0.00	10.00	0.00	4.17	
6. ineligibility	3.51	2.52	0.00	0.00	0.00	0.00	2.27	
7. other reasons	11.40	14.15	15.94	25.00	10.00	100.00	14.20	
(i) sub-total (rows 2–7)	79.82	66.35	47.83	50.00	70.00	100.00	66.48	
8. not recorded	6.14	9.12	13.04	6.25	10.00	0.00	8.90	
9. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	
no. of sample persons	(114)	(318)	(69)	(16)	(10)	(1)	(528)	
(b) base : unemployed persons with each type of attitude to registration								
1. registered	12.31	60.00	20.77	5.38	1.54	0.00	100.00	(130)
not registered due to								
2. ignorance	35.16	60.99	2.20	1.10	0.55	0.00	100.00	(182)
3. remoteness	0.00	40.00	40.00	20.00	0.00	0.00	100.00	(5)
4. lack of faith	10.91	60.00	20.00	1.82	7.27	0.00	100.00	(55)
5. indifference	18.18	54.55	22.73	0.00	4.54	0.00	100.00	(22)
6. ineligibility	33.33	66.67	0.00	0.00	0.00	0.00	100.00	(12)
7. other reasons	17.34	60.00	14.67	5.33	1.33	1.33	100.00	(75)
(i) sub-total (rows 2–7)	25.93	60.11	9.40	2.28	1.99	0.29	100.00	(351)
8. not recorded	14.89	61.70	19.15	2.13	2.13	0.00	100.00	(47)
9. total	21.59	60.23	13.07	3.03	1.89	0.19	100.00	(528)

number of sample towns=23 ; number of sample blocks=212 ; number of sample households=4,446

Appendix 1

COMPARISON OF TWO INDEPENDENT SUB-SAMPLE ESTIMATES

1. This survey on urban unemployment was the first of its kind to be undertaken by the National Sample Survey. It was in the nature of a try-out intended to reveal the various types of problems arising in surveys of this type. It was not expected to furnish estimates of high precision.

2. As with other surveys conducted by the NSS, the total sample was divided into two independent sub-samples, each capable of furnishing valid estimates. The number of sample towns in sub-sample I was 12 and it was 11 in sub-sample II. In some of the strata the number of sample towns represented in the two sub-samples was not the same. In the strata where the number of towns was not even, the number of towns finding a place in one sub-sample had to be one more or one less than in the other sub-sample. Nevertheless some idea of precision of the estimates can be had by comparing the two sub-sample estimates.

3. In the set of tables given below, a comparison has been made between the estimates obtained from the two sub-samples separately. A few important characteristics have been selected for comparison.

Table A.1 : Percentage distribution by age-groups.

Table A.2 : Percentage distribution by industrial status.

Table A.3 : Percentage distribution by economic status.

Table A.4 : Percentage distribution by general education.

Table A.5 : Percentage distribution by monthly household expenditure.

4. It will be observed from columns (2) and (3) in the annexed tables that the two sub-sample estimates are of the same order. In a good number of cases, however, the estimates may not be considered satisfactory if high precision is required. This is particularly true of estimates of cell frequencies in two-way distributions which have been generally given in the Statistical Tables.

5. The combined estimates shown in col (4) of the tables have not been derived by combining the two sub-sample estimates but they have been calculated directly from all the sample units taken together. This explains why, in some cases, the combined estimate does not lie between the two sub-sample estimates.

TABLE A.1 : PERCENTAGE DISTRIBUTION OF POPULATION BY AGE-GROUPS FOR TWO SUB-SAMPLES

age-groups (in years)	sub-sample I	sub-sample II	combined sub-sample estimates
(1)	(2)	(3)	(4)
1. 0— 6	20.10	19.20	19.61
2. 7—15	20.58	21.24	20.95
3. 16—17	3.69	3.59	3.61
4. 18—21	8.29	8.61	8.46
5. 22—26	10.10	9.74	9.94
6. 27—36	14.06	14.59	14.40
7. 37—46	9.67	10.00	9.91
8. 47—56	7.10	6.79	6.87
9. 57—61	2.75	2.62	2.66
10. 62 & above	3.66	3.62	3.59
11. total	100.00	100.00	100.00
no. of sample towns	12	11	23
no. of sample blocks	110	102	212
no. of sample households	2,394	2,052	4,446
no. of persons in the sample	10,812	9,593	20,405

TABLE A.2 : PERCENTAGE DISTRIBUTION OF POPULATION BY INDUSTRIAL STATUS FOR TWO SUB-SAMPLES

industrial status	sub-sample I	sub-sample II	combined sub-sample estimates
(1)	(2)	(3)	(4)
gainfully occupied persons:			
1. employers	0.36	0.56	0.47
2. employees	16.22	18.54	17.59
3. own-account workers	12.57	10.05	11.40
4. unpaid household labour	3.34	1.92	2.74
(i) sub-total :	32.49	31.07	32.20
unemployed:			
5. for the first time	0.98	1.48	1.16
6. not for the first time	1.19	1.76	1.43
(ii) sub-total :	2.17	3.24	2.59
7. family members engaged in domestic work only	22.95	21.71	22.28
8. rentiers and pensioners	0.65	0.67	0.61
9. beggars and persons living on charities and remittances	0.71	1.02	0.82
10. old and infirm persons, children and students	40.68	41.55	40.95
(iii) sub-total :	64.99	64.95	64.66
11. not recorded	0.35	0.74	0.55
12. total	100.00	100.00	100.00

TABLE A.3 : PERCENTAGE DISTRIBUTION OF POPULATION BY ECONOMIC STATUS FOR TWO SUB-SAMPLES

economic status	sub-sample I	sub-sample II	combined sub-sample estimates
(1)	(2)	(3)	(4)
1. earners	26.17	26.16	26.26
2. earning dependents	8.15	7.45	8.00
3. non-earning dependents	65.49	66.20	65.56
4. not recorded	0.19	0.19	0.18
5. total	100.00	100.00	100.00

TABLE A.4 : PERCENTAGE DISTRIBUTION OF POPULATION BY GENERAL EDUCATION FOR TWO SUB-SAMPLES

general education	sub-sample I	sub-sample II	combined sub-sample estimates
(1)	(2)	(3)	(4)
1. illiterate	64.69	48.47	57.36
2. literate but below matric	31.96	45.08	37.86
3. matric	2.15	4.33	3.17
4. intermediate	0.54	1.02	0.75
5. graduate and above	0.52	0.81	0.64
6. not recorded	0.14	0.29	0.22
7. total	100.00	100.00	100.00

TABLE A.5 : PERCENTAGE DISTRIBUTION OF POPULATION BY MONTHLY HOUSEHOLD EXPENDITURE CLASSES FOR TWO SUB-SAMPLES

expenditure class (in Rs.)	sub-sample I	sub-sample II	combined sub-sample estimates
(1)	(2)	(3)	(4)
1. up to 50	27.47	20.66	24.28
2. 51— 100	38.28	33.04	36.14
3. 101— 200	22.33	29.02	25.17
4. 201— 300	6.53	9.38	7.82
5. 301— 500	3.47	4.89	4.22
6. 501—1000	0.57	1.58	1.05
7. above 1000	0.25	0.92	0.52
8. not recorded	1.10	0.51	0.80
9. total:	100.00	100.00	100.00

APPENDIX 2

SCHEDULE OF INVESTIGATION

GOVERNMENT OF INDIA NATIONAL SAMPLE SURVEY : 1953-54

Preliminary Employment Survey : Urban Areas : September 1953

[1] identification		[2] general particular of household				[3] investigator: 1. name..................
1. zone-State		1. household size: male		9. h.h. industry occupation		....2. roll no.... 3. date survey..............
2. stratum		2. ,, female		10. land possessed (acre 00.00)		4. signature..............................
3. sample town		3. ,, total		11. religion		5. date schedule despatched................
4. sample block		4. no. of earners		12. mother tongue		[4] inspector: 1. name......................
5. fraction code		5. ,, earning dep.		13. informants' willingness		2. date of inspection ....
6. priority		6. non-earning dep.		14. informants' ability		3. date of scrutiny........................
7.s.h.h. no.		7. expenditure code(a)		15. name of head..................		4. signature...............................
		8. income code(b)				5. date schedule despatched ....

[5] demographic and employment details

as on date of survey......................									during month ended on.............									as on date of survey........			
																		if unemployed			
serial no.	relation to head	sex: m-1, f-2	age last birth day	marital status (c)	education (d)	migration (e)	why migrated (g)	economic status (h)	principal-1 subsidiary-2	industrial status (i)	employer: govt.-1 non govt.-2	duration of ind. status (f)	description of occupation	industry occupation code	intensity of employment (j)	why under-employed (k)	income code(b)	registration code(l)	why un-employed(k)	employment sought (m)	income code(b) (expected)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)

(a): *expenditure code:* 0-50 (0); 51-100 (1); 101-200 (2); 201-300 (3); 301-400 (4); 401-500 (5); 501-700 (6) 701-1000 (7), 1001-1500 (8); above 1500 (9).

(b): *income code:* 0-25 (01); 26-50 (02); 51-75 (11); 76-100 (12); 101-150 (21); 151-200 (22); 201-250 (31); 251-300 (32); 301-350 (41); 351-400 (42); 401-500 (50); 501-700 (60); 701-1000 (70);1001-1500 (80); above 1500(90).

(c): *marital status:* single-1, married-2, widowed-3, divorced-4, separated-5.

(d): *education code:* 2 digited—*left digit:* illiterate (0), literate but below primary (1), primary (2), middle (3), matric (4), intermediate (5), graduate-arts & commerce (6), graduate—science (7), postgraduate—arts & commerce (8), post-graduate—science (9).

right digit: no technical or professional qualification (o), professional or technical skill without any degree or equivalent diploma (1), typing, stenography diploma or certificate holder (2), nursing, midwifery and pharmacy diploma or certificate holders (3), degree or equivalent diploma holder in engineering (4), agriculture (5), medicine (6), banking, insurance, audit, accounts (7), law (8), other technology or profession (9).

2

[6] employment position a year ago from date of survey......

serial no. in bl. 5	weeks unemployed during last year	economic status (h)	principal-1 subsidiary-2	industrial status (i)	employer: govt.-1 non-govt-2	duration of ind. status (f)	description of occupation	industry occupation code	intensity of employment (j)	why under-employed (k)	income code (b)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)

[7] time lag between completion of education & first employment of graduates & under-graduates

serial no.	serial no. in bl. 5	years since education competed (f)	interval: first employment (f)	description of occupation	industry occupation code	income code (b)
(1)	(2)	(3)	(4)	(5)	(6)	(7)

[8] total cash receipts during last month ended on ......

serial no.	type of receipts	amount (Rs.)
(1)	(2)	(3)
1.	earned income	
2.	rents	
3.	pensions	
4.	loans taken	
5.	sale of assets	
6.	past savings	
7.	charities	
8.	remittances	
9.	others	
10.	total	

list of code nos. (contd):

(e): *migration code—2 digited*

left digit: local (0), temporary visitor (1), migrant from rural area of same state (2), urban area of same state (3), rural area of different state (4), urban area of diffe rent state (6), East Pakistan as D.P. (6), West Pakistan as D.P. (7), other migrants (8).
right digit: use duration code.

(f) *duration code:* nil (0), 3months (1), 6 months (2), 1 year (3), 2 years (4), 4 years (5), 6 years (6), 10 years (7), 15 years (8), above 15 years (9).

(g) *why migrated:* unemployment (1), partial unemployment (2), occupation not congenial (3), further studies not possible (4), other reasons (5), does not arise (6).

(h) *economic status:* earner ((1); earning dependent (2); nonearning dependent (3).

(i) *industrial status:* employer (1), employee (2), own account (3), unpaid h.h. labour (4), unemployed seeking employment for the first time (5), other unemployed seeking employment (i.e. excluding 5) (6), family member engaged in domestic work only (7), rentiers, pensioners (8), beggars and persons living on charities and remittances (9), old, infirm, child, student (0).

(j) *intensity of employment:*no work (0), less than quarter (1), quarter (2), half (3), three quaters (4), full (5).

(k): *why underemployed:* disability to accept work for-illness (0), domestic reasons (1) ; fall in production due to-lack of materials and equipments (2), fall in demand (3), slack season (4), policy, political & other reasons not known (5); industrial disputes (6), other reasons (7).

(l) *registration code:* registered (1) not registered due to—ignorance (2), remoteness (3), lack of faith (4), indifference (5), ineligibility (6), other reasons (7).

(m) *employment sought: seeking to start own business:* agricultural production (1), animal husbandry(2), other manufacture and handicrafts (3). transport service (4), trade (5), professions and financial transactions (6).

seeking to obtain employment as employee: unskilled manual work (8). skilled manual work (artisan) (9), administrative, executive or clerical work, including accounts and typing etc. (10), nonmanual work —technical and technological (11), nonmanual work—medical, surgical, etc. (12), nonmanual work—legal, teaching, writing and the like (13), nonmanual work fine arts, including production of cinema films, theatrical performances etc. (14), nonmanual work other kinds (15).

E. P.—781-14.8.53.—10,000

SUR LA CONVERGENCE STOCHASTIQUE AU SENS DE CESARO ET SUR DES DIFFÉRENCES IMPORTANTES ENTRE LA CONVERGENCE PRESQUE CERTAINE ET LES CONVERGENCES EN PROBABILITÉ ET PRESQUE COMPLETES

Par D. DUGUE

de l'Institut de Statistique de l'Université de Paris

Le but de cet article est de rattacher les réciproques des théorèmes connus sur le comportement de la moyenne de variables aléatoires à des théorèmes sur le comportement de la plus petite et de la plus grande de ces valeurs. Les résultats que nous donnons sont souvent valables en dehors du cas de l'échantillon. Nous utiliserons des démonstrations que nous avons exposés dans un article (D. Dugué. Eléments stochastiques limités) (Bulletin de l' Institut International de Statistique 28^e session (Rome)) qui sera désigné par A, et dans un article qui va paraître dans la même revue sous le titre : Sur les théorèmes limités du calcul des probabilités (B), ainsi qu'une inégalité fondamentale de M. Paul Levy (voir par exemple Theorie de l'addition des variables aléatoires chapitre VI, paragraphe 44). En particulier la définition de la convergence presque complète est donnée dans A.

1. LE COMPORTEMENT DE LA MOYENNE DE VARIABLES ALÉATOIRES ET LE COMPORTEMENT DE LA PLUS PETITE ET DE LA PLUS GRANDE

Nous allons tout d'abord établir le théoréme suivant.

Theoreme 1.1 : $X_1, X_2, \dots, X_n, \dots,$ *étant une suite de variables aléatoires et* M_n *et* m_n *respectivement la plus grande et la plus petite valeur des n premiers résultats, il est nécessaire pour que la suite des variables* $Y_n = \frac{S_n}{n} = \frac{X_1, \dots, +X_n}{n}$ *converge presque sûrement vers une quantité certaine ou aléatoire que* $\frac{M_n}{n}$ *et* $\frac{m_n}{n}$ *tendent presque sûrement vers* 0.

Soit $x_1, x_2, \dots, x_n, \dots,$ une suite de réalisations des variables aléatoires $X_1, X_2, \dots, X_n, \dots$. Si y_n tend vers une limite (aléatoire ou certaine) il est connu que $\frac{1}{n}$ [la plus grande valeur des n premiers résultats] et $\frac{1}{n}$ [la plus petite valeur des n premiers résultats] tendent vers 0. Le premier évenement (existence d'une limite pour y_n) entraînant le second (limite nulle pour $\frac{M_n}{n}$ et $\frac{m_n}{n}$) le second a une probabilité supérieure ou égale à celle des premiers (c'est-à-dire 1.)

Il est donc presque certain que $\frac{M_n}{n}$ et $\frac{m_n}{n}$ tende vers 0. Remarquons que le théorème ne fait aucune hypothése sur la dépendance ou l'indépendance des variables $X_1, \ldots, X_n \ldots$.

Dans le cas de l'échantillon ($X_1, \ldots, X_n, \ldots$ indépendants et ayant la même loi de probabilité) il est nécessaire pour que $\frac{M_n}{n}$ et $\frac{m_n}{n}$ tendent presque sûrement vers 0 que $E(|X|)$ existe; on retrouve ainsi comme ces cas particuliers un résultat déjà connu (voir par exemple D. Dugué, C. R. Ac. Sc. Paris 234 p. 1837).

On a de même le théorème I bis.

Theoreme 1.2: $X_1, X_2, \ldots, X$, ... *étant une suite de variables aléatoires pour que* Y_n *soit presque sûrement borné en valeur absolue par un nombre certain* M *il est nécessaire que* $\frac{M_n}{n}$ *et* $\frac{m_n}{n}$ *soient presque sûrement bornés en valeur absolue par* $2M$.

D'une manière générale on dit qu'une suite aléatoire est presque sûrement bornée par le nombre certain M s'il y a une probabilité égale à 1 pour qu'il $n'y$ ait qu'un nombre fini de résultats supérieurs à M.

Si l'on a presque sûrement $\left|\frac{X_1+, \ldots ,+X_n}{n}\right| < M$ on aura presque sûrement $\left|\frac{X_1+, \ldots, +X_n}{n+1}\right| < \frac{n}{n+1} M$ et par suite $\left|\frac{X_{n+1}}{n+1}\right| < 2M$. Cela entraine exactement comme dans le cas du théorème I que $\frac{M_n}{n}$ et $\frac{m_n}{n}$ soient presque sûrement bornés en valeur absolue par $2M$.

Dans le cas ou $X_1, \ldots, X_n$ forment un échantillon, on sait (voir A) que si $\frac{M_n}{n}$ et $\frac{m_n}{n}$ sont presque sûrement bornés, ils tendent presque sûrement vers 0 et par suite que $\frac{X_1+ \cdots +X_n}{n}$ tend presque sûrement vers une limite (qui est égale à $E(X)$). Dans le cas de l'échantillon $\frac{X_1+ \cdots +X_n}{n}$ ne peut pas être presque sûrement borné, sans tendre presque sûrement vers une limite

On a de même.

Theoreme 1.3: $X_1, \ldots, X_n$ *formant un échantillon* Y_n *ne peut pas être presque sûrement stable sans que* Y_n *tende presque sûrement vers une limite.*

En effet si Y_n est presque sûrement stable il existe une suite de nombres certains $a_1, \ldots, a_n, \ldots$ tels que $Y_n - a_n$ tende presque sûrement vers 0.

On en deduit que $\frac{X_{n+1}}{n+1} - a_{n+1} + \frac{n}{n+1} a_n$ doit tendre presque sûrement vers 0.

Par conséquent $\frac{X_{n+1}}{n+1} - \left(a_{n+1} - \frac{n}{n+1} a_n\right)$ doit tendre en probabilité vers 0. Les X_n obéissant à la même loi de probabilité $\frac{X_{n+1}}{n+1}$ tend en probabilité vers 0. On en déduit que $a_{n+1} - \frac{n}{n+1} a_n$ doit tendre vers 0. Ce qui a comme conséquence que $\frac{X_{n+1}}{n+1}$ tend presque sûrement vers 0. Or si dans un échantillon $\frac{X_n}{n}$ tend presque sûrement vers 0, Y_n tend presque sûrement vers une limite égale à $E(X)$ ce qui établit le théorème 1.3.

On peut généraliser cet énoncé en supposant seuelement que $X_1, \ldots, X_n, \ldots$ forment un ensemble de variables aléatoires borné en loi. (pour la définition d'ensembles bornés en loi voir B).

Dans ces conditions $\frac{X_n}{n}$ tend en probabilité vers 0. On en déduit encore que $a_{n+1} - \frac{n}{n+1} a_n$ tend vers 0; ce qui entraîne encore que $\frac{X_{n+1}}{n+1}$ tend presque sûrement vers 0 et par suite que $\frac{M_n}{n}$ et $\frac{m_n}{n}$ tendent presque sûrement vers zéro.

On a donc.

Theoreme 1.4: $X_1, \ldots, X_n, \ldots$ *étant des variables aléatoires bornés en loi si* Y_n *est presque sûrement stable il est nécessaire que* $\frac{M_n}{n}$ *et* $\frac{m_n}{n}$ *tendent presque sûrement vers* 0.

2. LE COMPORTEMENT EN PROBABILITÉ ET LE COMPORTEMENT PRESQUE COMPLET DES QUANTITES Y_n

Il est curieux de remarquer qu'il *y* a plus d'analogie entre les théorèmes qui concernent ces deux sortes de convergence qu'avec ceux concernant la convergence presque certaine.

Tout d'abord remarquons que l'indépendance des variables est ici essentielle. On peut avoir une suite de variables *non indépendantes*, telles que Y_n tende en probabilité vers une limite alors que $\frac{M_n}{n}$ et $\frac{m_n}{n}$ ne tendent pas en probabilité vers 0. Et il en est de même presque complètement.

En effet considérons la suite :

$$X_1, X_2 = -X_1, X_3, X_4 = -X_3, \ldots \quad X_{2n-1}, X_{2n} = -X_{2n-1}, \ldots,$$

les variables $X_1, X_3, \ldots, X_{2n-1}, \ldots$ étant indépendantes et obéissant à la même loi de probabilité $F(x)$ (autrement dit formant un échantillon).

class, then we can expect that we will not be able to make as precise statements as if we had applied a technique specially designed for this latter class of models. If we insist in demanding moderate or high precision in our estimates, quite long time series will be required or alternatively a fair number of shorter series. In connection with this it may be pointed out that most of the recent work is of a large sample character.

As we have seen the three groups of models complement each other and correspond to different practical problems. Although much progress has been made in time series analysis, it is evident that much has to be done before the theory can be said to be in a satisfactory shape.

The present fairly general approach to time series analysis is somewhat hampered by the lack of suitable families of stochastic processes defined to fit the demand from the applications. Therefore the most important need is perhaps that of more realistic and flexible probability models. To do this we have to rely on the applications : close contact with the subject sciences will help us to construct such classes of stochastic processes that are well fitted for practical use. The success of some of the new techniques mentioned in Section 4 is undoubtedly due to the fact that they are founded on models with a sound deductive background. Instead of just trying to fit a model mechanically to the data, one should seek one based on a more penetrating analysis. The class of probability structures ought to correspond to the knowledge we already have of the phenomena under investigation. Causal concepts may be helpful for this construction. An illuminating discussion of this and related topics can be found in H. Wold (1952).

It seems likely that also in future work in time series analysis we will have to be satisfied with approximate methods. I would be surprised if it showed possible to obtain methods of inference valid exactly for small samples under such general conditions that we are interested in. In the light of what has been said earlier, this need not depress any except the perfectionists, since we will need fairly large samples anyway to get sufficient resolving power in the analysis. The modern tendency among theoretical statisticians of to-day towards abstract methods, valuable and necessary also in time series analysis, has resulted in that problems with a more practical flavour have been neglected to some extent. The extensive formal manipulations may scare the theoretician and offend his aesthetic sense, but there are a lot of interesting, difficult and tempting questions involved, usually of an analytic character, also when approximate methods should be designed. This is so at least if one has ambitions to get logically sound and mathematically proved results.

This brings me to another point which I consider important. It has been argued that, as statistics is an entirely applied science, it is not necessary to have completely rigorous derivations. Now the statistician usually takes the attitude that one of this jobs (and perhaps the main one) is to spread the use of sound techniques in research when statistics is used in practice. We all know how easy it is to make an error in the planning and analysis of statistical investigations. The statistician should see that this does not happen, and he is then helped by his experience and by statistical theory. If this should work it is necessary that the theory is correct. Now, in many branches of statistics the mathematical difficulties are negligible compared to the questions of application and interpretation of the method. If the algebra and calculus involved is not used in a quite rigorous way, there still is not very great danger of committing a grave error, if the method is based upon a sound intuitive argument. However as the mathematical complexity increases, as has happened in time series analysis and which is likely to go on, then one has to be more careful with the mathematics. There are examples in time series analysis where errors of this kind have led to misleading or erroneous results. I do not want to underestimate the role of intuition in scientific work, but we should be careful before offering a new statistical technique, so that we know if and when it works. It is only a simple act of intellectual honesty to apply the same norms on the logical basis for theoretical work as when we criticize the subject scientist for his defect statistical reasonings. We have all met the holier-than-thou attitude among some statisticians towards their colleagues in applied fields. Furthermore a correct mathematical proof interpreted in the right manner tells us a good deal about the range of the methods: proofs are not a luxuary but can and should be used to tell us practical things. Also let us note that the practical statistician does not have the same possibilities of empirically verifying or falsifying the validity of his method as does *e.g.* the physicist. The only way in which the statistician can do this is by artificial sampling experiments. This is quite costly and cumbersome to do in time series analysis, at least that is my experience.

I would like to discuss briefly another topic of considerable importance, how to use time series for the purpose of prediction ? The theory of prediction due to Wiener-Kolmogorov (see Wiener 1949) is well-known. It is designed for the case when we have observed a stationary process (say the one we have denoted by $y(t)$) up till some time t. We want to predict the value of the process at some future time point, $y(t+h)$. This beautiful theory has many important applications, but in many practical problems it cannot be used since either $y(t)$ is not stationary or because there is an unknown systematic component $m(t)$ superimposed on the observations. If we do not have a good idea of the underlying mechanism, we may be unable to describe $m(t)$ closely enough. This is certainly a case where overenthusiastic use of existing tools of time series analysis might lead to disastrous results.

Situations like this exist for long range forecasting in fields like economics, demography and others. It is clear that the solution does not lie only in a refined statistical treatment but more information on the functioning of the generating mechanism must be collected so that the structure of $m(t)$ and $y(t)$ can be described in more precise terms. Statistics does not work miracles just because there is a superstitious belief in it.

6. CONCLUSIONS

Summing up I would put the emphasis on the following items :

(i) Keeping in touch with the practical demands the time series analysts and the probabilists together must invent new classes of stochastic models, based on more realistic assumptions about the underlying mechanism.

(ii) Starting from the appropriate models statistical methods should be developped, which are applicable for at least large and moderate sample size. The validity of the approximations should be examined. Attention should be given to the need for simpler methods.

(iii) The theoretical investigations should be based on a rigorous logical and mathematical foundation.

I do not want to give the impression that what I have said here represents the common and only opinion among time series analysts. Some of the topics discussed are quite controversial, but I would be glad if this paper could stimulate to a discussion of these questions.

REFERENCES

ANDERSON, R. L. (1942): Distribution of the serial correlation coefficient. *Ann. Math. Stat.*, **13**, 1-13.

BARTLETT, M. S. (1946): On the theoretical specification and sampling properties of autocorrelated time-series. *J. Roy. Stat. Soc.*, Supplement, **8**, 27-41.

——— (1950): Periodogram analysis and continuous spectra. *Biometrica.*, **37**, 1-16.

——— (1955): *An Introduction to Stochastic Processes*, Cambridge University Press, London.

CRAMÉR, H. (1942): On harmonic analysis in certain functional spaces. *Arkiv Mat. Astr. Fys.*, **28B**, 7.

DOOB, J. L. (1953): *Stochastic Processes*, John Wiley and Sons, New York.

FISHER, R. A. (1929): Test of significance in harmonic analysis. *Proc. Roy. Soc.*, *A*, **125**, 54-59.

GRENANDER, U. (1950): Stochastic processes and statistical inference. *Arkiv Mat.*, **1**, 195-277.

——— (1951): On empirical spectral analysis of stochastic processes. *Arkiv Mat.*, **1**, 503-531.

——— (1954): On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Stat.*, **25**, 252-272.

——— and M. ROSENBLATT (1953): Statistical spectral analysis of time series arising from stationary stochastic processes. *Ann. Math. Stat.*, **24**, 537-558.

KOOPMANS, T. (1942): Serial correlation and quadratic forms in normal variables. *Ann. Math. Stat.*, **13**, 14-33

MANN, H. B. (1953): *Introduction to the Theory of Stochastic Processes Depending on a Continuous Parameter*, National Bureau of Standards U.S.A.

TUKEY, J. W. : *Measuring Noise Color*, unpublished manuscript.

WIENER, N. (1949): *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*, John Wiley and Sons, New York.

WOLD, H. (1938): *A study in the Analysis of Stationary Time Series*, Almqvist and Wiksell, Stockholm.

——— (1949): A large-sample test for moving averages. *J. Roy. Stat. Soc.*, *B.*, **11**, 297-305.

——— (1952): *Demand analysis. A study in Econometrics*, John Wiley and Sons, New York.

YULE, G. U. (1927): On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Trans. Roy. Soc.*, *A.*, **226** 267-298.

Paper received : February, 1956.

THE USE OF SAMPLE RANGE IN ESTIMATING THE STANDARD DEVIATION OR THE VARIANCE OF ANY POPULATION

By MOTOSABURO MASUYAMA

Meteorological Research Institute, Tokyo
and
Indian Statistical Institute, Calcutta

1. Introduction

In statistical quality control and sometimes in the analysis of large-scale sample surveys, the sample range is used to estimate the standard deviation or the variance in the population. This requires the use of values of the ratio of the expectation of the sample range to the standard deviation as a function of the sample size which has been tabulated for a normal population and is used extensively in industrial practice. For a general continuous distribution, Plackett (1947) gave an upper bound for this ratio and at the same time gave a discontinuous distribution in which this ratio is nearly zero. Tsumura (1952) computed this ratio for rectangular, triangular, exponential and normal distributions and found that if the sample size is small, these values are nearly equal, and fairly close to the upper bound obtained by Plackett.

A problem that arises is this. Can an estimate of the standard deviation or the variance be obtained by using a constant multiple of the range or its square, with the same multiplier for all distributions? In this paper bounds are derived for the ratio of the expectation of the sample range to the population standard deviation and that of the square of the sample range to the population variance. In the opinion of the author, these bounds are quite close, specially if the sample is small, and the harmonic mean of these bounds is suggested as the appropriate multiplier.

2. Estimation of the standard deviation

Let u denote the standard deviation and R the range in a sample of size n. Neglecting the trivial case $R = 0$, it is easy to show that

$$U = \frac{u}{R_{\max}} = \begin{cases} 1/2\sqrt{n/(n-1)} & \text{if } n \text{ is even} \\ 1/2\sqrt{(n+1)/n} & \text{if } n \text{ is odd} \end{cases} \qquad \dots \quad (2.1)$$

$$L = \frac{u}{R_{\min}} = \sqrt{1/2(n-1)}. \qquad \dots \quad (2.2)$$

If we now define coefficients b and d by

$$E(u) = b\sigma \qquad \dots \quad (2.3)$$

$$E(R) = d\sigma \qquad \dots \quad (2.4)$$

$A \int^{+\infty} \frac{dx}{x^2 \log |x|} = 1$ (nous allons préciser plus tard les intervalles où cette densité est effectivement 0 ou $\frac{A}{x^2 \log |x|}$). On a : $\int_u^{+\infty} \frac{dx}{x^2 \log |x|} < \frac{1}{u \log u}$ et par conséquent la première condition est remplie. D'autre part $\int^{+\infty} \frac{x dx}{x^2 \log |x|}$ est divergente.

Prenons la densité de probabilité égale à 0 sur le segment $-2, +2$ prenons la égale à $\frac{A}{x^2 \log |x|}$ de $+2$ à x_1 et nulle de -2 à $-x_1$. x_1 étant déterminé par la condition $\int_2^{x_1} \frac{A x dx}{x^2 \log |x|} = M$, qui est possible à cause de la divergence de l'intégrale. Prenons la densité égale à 0 de x_1 à x_2 et égale à $\frac{A}{x^2 \log |x|}$ de $-x_1$ à $-x_2$, x_2 étant déterminée par $\int_{-x_2}^{-x_1} \frac{A x dx}{x^2 \log |x|} = \alpha_1$ avec $\alpha_1 = -2M$; et ainsi de suite, on voit que l'intégrale au sens de Cauchy $\int_{-n}^{+n} \frac{A x dx}{x^2 \log |x|}$ va osciller entre $-M$ et M (à condition que la valeur absolue de toutes les quantités analogues à α_1 soit égale à $2M$, ces quantités étant alternativement positives ou négatives) et ne tendra donc vers aucune limite tout en restant bornée.

Examinons maintenant le cas de la stabilité.

Theoreme 2.4: *Si $X_1, \ldots, X_n, \ldots$ forment un échantillon et si Y_n est stable en probabilité* (*ou presque complètement*) $\frac{M_n}{n}$ *et* $\frac{m_n}{n}$ *tendent en probabilité* (*ou presque complètement*) *vers* 0.

Si $\frac{S_n}{n}$ est stable en probabilité il existe une suite $a_1, \ldots, a_n, \ldots$ telle que $\frac{S_n}{n} - a_n$ tende en probabilité vers 0. D'après le paragraphe I, on doit avoir $a_{n+1} - \frac{n}{n+1} a_n$ tendant vers 0.

$\frac{S_m - S_q}{m} - a_m + a_q \frac{q}{m} = \left(\frac{S_m}{m} - a_m \right) - \left(\frac{S_q}{q} - a_q \right) \frac{q}{m}$. Donc $\frac{S_m - S_q}{m} - a_m + a_q \frac{q}{m}$ tend en probabilite vers 0 quand m augmente indefiniment (avec $m > q$).

On peut donc en prenant n assez grand trouver des nombres $b_1, \dots b_n$ aussi voisins de 0 qu'on le desire, tels que l'intervalle ferme de concentration maximum pour $1 = \frac{\epsilon}{2}$ contienne l'origine comme point interieur pour les variables suivantes :

$$\frac{S_n - S_1}{n} - a_n + a_1 \frac{1}{n} - b_n + b_1$$

$$\frac{S_n - S_2}{n} - a_n + a_2 \frac{2}{n} - b_n + b_2.$$

Par consequent en vertu de l'inegalite de M. Paul Levy, la probabilité pour que toutes les inegalités

$$|S_1 - a_1 - nb_1| \leqslant n\epsilon, \dots, |S_n - na_n - nb_n| \leqslant n\epsilon$$

soient satisfaites est superieure $1 - \dfrac{1 - Q_{s_n}\left(n \frac{\epsilon}{2}\right)}{Q_{s_n}\left(n \frac{\epsilon}{2}\right)}$

Le fait que toutes les inegalités soient satisfaites entraine que les inegalites

$$\left|\frac{X_1}{n} - \frac{a_1}{n} - b_1\right| \leqslant \epsilon$$

$$\left|\frac{X_2}{n} - \left(a_2 - \frac{a_1}{2}\right)\frac{2}{n} - (b_2 - b_1)\right| \leqslant 2\epsilon$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\left|\frac{X_{p+1}}{n} - \left(a_{p+1} - \frac{p}{p+1}\, a_p\right)\frac{p+1}{n} - (b_{p+1} - b_p)\right| \leqslant 2\epsilon$$

$$\left|\frac{X_n}{n} - \left(a_n - \frac{n-1}{n}\, a_{n-1}\right)\frac{n}{n} - (b_n - b_{n-1}\right| \leqslant 2\epsilon$$

sont toutes satisfaites et comme pour n assez grand $b_1, \dots, b_n$ peuvent etre choisis arbitrairement voisins de o et que $a_{p+1} - \frac{p}{p+1}\, a_p$, tend vers 0 quand p augmente indefiniment on en deduit comme plus haut que :

$$\text{prob}\ \left[\frac{M_n}{n} \leqslant 3\epsilon\right] > 1 - \frac{1 - Q_{s_n}\left(n \frac{\epsilon}{2}\right)}{Q_{s_n}\left(n \frac{\epsilon}{2}\right)}$$

et la même inegalite pour m_n. Cela entraine que $\frac{M_n}{n}$ et $\frac{m_n}{n}$ tendent en probabilité (on presque complètement) vers 0 quand $\frac{S_n}{n}$ est stable en probabilité (ou presque complètement).

Car quand $\frac{S_n}{n}$ est stable en probabilité $Q_{s_n}\left(n\,\frac{\epsilon}{2}\right)$ tend vers 1 quel que soit $\epsilon > 0$ quand n augmente indéfiniment et quand $\frac{S_n}{n}$ est stable presque complétement $1-Q_{s_n}\left(n\,\frac{\epsilon}{2}\right)$ est le terme général d'une série convergente.

Corollaire; *Dans le cas de l'échantillon la moyenne* $Y_n = \frac{X_1+\ldots+X_n}{n}$ *ne peut être presque complètement stable sans tendre presque complètement vers une limite.*

Car le fait que la moyenne soit stable entrainant que $\frac{M_n}{n}$ tende vers 0 (le tout presque complètement) $E(X^2)$ existe ce qui a pour conséquence que Y_n tende vers une limite.

Nous venons de voir qu'il n'en est pas de meme en probabilité.

3. CONVERGENCE EN PROBABILITÉ VER 0 SANS CONVERGER EN PROBABILITÉ AU SENS DE CESARO

Nous allons maintent donner des exemples des faits suivanti :

Une suite de variables aléatoires $X_1, X_2, \ldots, X_n, \ldots$ peut tendre presque complètement vers 0 sans tendre vers une limite presque compléte au sens de Cesaro.

Par contre, si une suite de variables aléatoires $X_1, X_2, \ldots, X_n, \ldots$ tend presque sûrement vers 0, elle tend presque sûrement vers 0 au sens de Cesaro. Cette différence dans le comportement provient du fait que la limite en probabilité ou presque compléte est une propriété marginale, alors que la limite presque sure est une propriété globale.

Si une suite de réalisations tend vers 0, cette suite de réalisations tend aussi vers 0 au sens de Cesaro et par conséquent si les suites de réalisations tendant vers 0 ont une probabilité unité les suites ayant 0 pour limite au sens de Cesaro ont aussi une probabilité unité, ce qui établit le théoréme concernant la limite presque certaine.

Voici un exemple d'un cas ou X_n tend vers 0 en probabilité sans converger en probabilité au sens de Cesaro

Prenons les X_n indépendants avec :

$$F_{X_n}(x) = 1-\frac{1}{x+n} \text{ pour } x > 0 \text{ et } F_{X_n}(x) = 0 \text{ pour } x \leqslant 0$$

X_n est toujours positif et converge en probabilité vers 0 puisque :

$$1-F_{X_n}(\epsilon) = \frac{1}{\epsilon+n} \text{ qui tend vers 0 avec } \frac{1}{n} \text{ quel que soit } \epsilon > 0$$

D'autre part pour que X_n converge en probabilité au sens de Cesaro, il est nécessaire que $\frac{M_n}{n}$ converge en probabilité vers 0. Or la probabilité pour que $\frac{M_n}{n}$ soit inférieure à ϵ est :

$$\left(1-\frac{1}{n\epsilon+1}\right)\left(1-\frac{1}{n\epsilon+2}\right)\dots\left(1-\frac{1}{n\epsilon+n}\right) \text{ qui est inférieur à } \left(1-\frac{1}{n\epsilon+n}\right)^n \text{ et}$$

ne tend pas vers 1 quand n augmente indéfiniment.

Le cas de la convergence presque complete sera donné par la suite de variables indépendantes X_n positives avec:

$$F_{X_n}(x) = 1-\frac{1}{(x+n)\log^2 2n} \text{ pour } x > 0 \text{ et } F_X(x) = 0 \text{ pour } x \leqslant 0.$$

La suite converge presque complètement puisque :

$$1-F_{X_n}(\epsilon) = \frac{1}{(n+\epsilon)\log^2 2n} < \frac{1}{n\log^2 2n}$$

est le terme général d'une série convergente quel que soit ϵ positif.

La probabilité que $\frac{M_n}{n}$ soit supérieure à ϵ sera ici :

$$1-\left(1-\frac{1}{(n\epsilon+1)\log^2 2}\right)\left(1-\frac{1}{(n\epsilon+2)\log^2 4}\right)\dots\left(1-\frac{1}{(n\epsilon+n)\log^2 2n}\right)$$

qui sera donc supérieur à :

$$1-\left(1-\frac{1}{(n\epsilon+n)\log^2 2n}\right)^n \text{ équivalent a } \frac{1}{(\epsilon+1)\log^2 2n}\Bigg)$$

Par conséquent $\frac{M_n}{n}$ est presque conplètement supérieur à ϵ quel que soit ϵ et X_n ne tend pas presque complétement vers 0 au sens de Cesaro.

Si une suite converge presque sûrement vers 0 (et a fortiori si elle converge presque completèment) comme elle converge presque sûrement au sens de Cesaro vers 0 et que la convergence presque sûre entraine la convergence en probabilité, elle converge en probabilité au sens de Cesaro vers 0.

On voit donc les differences qu'il y a entre ces différentes limites.

Les deux résultats obtenus montrent que la convergence en probabilité et la convergence presque compléte ne sont pas compatibles avec l'existence d'une norme obéissant aux postulats habituels de la norme

a) $\| X, X \| = 0,$ b) $\| X, Z \| \leqslant \| X, Y \| + \| Y, Z \|,$

c) $\| \alpha X, \alpha Y \| = \alpha \| X, Y \|$

En effet si dans un espace normé X_n tend vers 0 on en déduira :

$$\| X_n, \ 0 \| \to 0$$

Et :

$$\left\| \frac{X_1+,\ldots,+X_n}{n}, \ 0 \right\| \leqslant \frac{\| X_1, 0 \| + \| X_2, 0 \| +,\ldots,+ \| X_n, 0 \|}{n}$$

qui tend vers 0 si $\| X_n, \ 0 \|$ tend vers 0.

Donc dans un espace normable la convergence ordinaire entraîne la convergenceau sens de Cesaro.

M. Fréchet a établi (cf. Généralités sur les probabilités : eléments aléatoires) que la convergence presque sûre n'est pas compatible avec l'existence d'une distance. Par contre il a montré que l'existence d'une distance est compatible avec la convergence en probabilité.

English summary

In this paper the author gives theorems binding the convergence (on different meanings, in probability, almost certainly, almost completely) of the mean of a sample and the same convergence towards 0 of $\frac{M_n}{n}$ and $\frac{m_n}{n}$, M_n and m_n being the maximum and minimum of the sample and n its size. He uses a Paul Levy's inequality.

"Almost complete convergence" in this paper is the same as "complete convergence" introduced by P. L. Hsu and H. E. Robbins (Proc. Nat. of Sciences). The author added the word "almost" because it seemed to him more adequate. A set of random variable $X \ldots X_n$, is said to converge completely (Hsu and Robbins' notion) and "almost completely" (Dugue's notion) towards 0 if a set of *independent* variables having the same marginal laws as X_1, X_{n1}—converges almost surely towards 0.

By "borné en loi" the author means that the inferior bound (for $X > 0$) tends towards 1 and the superior bound (for $X < 0$) tends towards 0 when $/x/$ increases indefinitely.

Paper received : January, 1956.

MAXIMUM LIKELIHOOD ESTIMATION FOR THE MULTINOMIAL DISTRIBUTION

By C. RADHAKRISHNA RAO

Indian Statistical Institute, Calcutta

1. INTRODUCTION

In an earlier paper (Rao 1953), the author considered minimum Chi-Square [m.c.] estimates and derived some of their large sample properties. Using the same techniques, it is possible to derive similar properties of maximum likelihood [m.l.] estimates, and these properties seem to be worth recording. For instance, it is not explicitly made clear in literature what analytical properties are possessed by m.l. estimates, under the usual regularity conditions (see section 4 of this paper) on the hypothetical cell frequencies as functions of unknown parameters. In fact, many probability statements concerning the m.l. estimates are direct consequences of the continuity and differentiability of the m.l. estimate as a function of observed relative frequencies.

Wald (1949) proved, under some regularity conditions, that the m.l. estimate is consistent while Huzurbazar (1949) showed, under Cramer's regularity conditions (Cramer, 1946), that with probability tending to unity a consistent root of the m.l. equation provides a local maximum of the likelihood. It was not known whether at such a root the likelihood attains an absolute maximum. It is shown in this paper that in the case of the multinomial distribution, a m.l. estimate is, with probability 1, a root of the likelihood equation, and provides a maximum of the likelihood when the parameter is restricted to the roots. We thus have a complete theory of the method of m.l. for the multinomial distribution. No assumption is made about the existence of the m.l. estimate, but the existence and uniqueness are deduced as a consequence of the regularity conditions.

To make the arguments free from unnecessary complications only the one parameter case is considered. The additional complication in the proof for the multiparameter case is in establishing the existence of the roots of the likelihood equation. Once this is done, the rest of the argument is the same.

In the development of this paper, the following scheme is adopted. First, the consistency of the m.l. estimates of the hypothetical frequencies of the multinomial distribution is established without any regularity conditions whatsoever. Second, the consistency of the m.l. estimate of a parameter occurring in the specification of the hypothetical frequencies is established under a very natural restriction on the parameter, Third, the m.l. estimate is identified (with probability 1 as the sample size tends to infinity) as *the* root of the likelihood equation providing the maximum of the likelihood with the parameter restricted to the roots, and its analytical properties discussed.

2. DEFINITIONS AND PRELIMINARY LEMMAS

The hypothetical frequencies in k classes are represented by $\pi_1(\theta), \ldots, \pi_k(\theta)$ and the observed relative frequencies by $p_1, \ldots, p_k$.

Definition 1: The likelihood of the hypothetical frequencies $\pi_1, \ldots, \pi_k$ given the observed relative frequencies $p_1, \ldots, p_k$ in a sample of size n is proportional to

$$L(\pi) = \pi_1^{np_1} \ldots \pi_k^{np_k} \qquad \ldots \quad (2.1)$$

Definition 2.1: The m.l. estimate of the multinomial distribution is a set of values $\pi_1, \ldots, \pi_k$ (if it exists) belonging to a given admissible class A of distributions and for which the likelihood $L(\pi)$ or the equivalent expression

$$p_1 \log \pi_1 + \ldots + p_k \log \pi_k \qquad \ldots \quad (2.2)$$

is a maximum when π is restricted to A.

Definition 2.2: The m.l. estimate of a parameter θ occurring in the specification of the hypothetical frequencies is a value θ, if one exists in the admissible set of values of θ, for which

$$p_1 \log \pi_1(\theta) + \ldots + p_k \log \pi_k(\theta) \qquad \ldots \quad (2.3)$$

is a maximum.

Definition 3: The m.l. equation is obtained by equating the derivative of (2.1) to zero, i.e.,

$$\frac{p_1}{\pi_1} \frac{d\pi_1}{d\theta} + \ldots + \frac{p_k}{\pi_k} \frac{d\pi_k}{d\theta} = 0 \qquad \ldots \quad (2.4)$$

Of course, the differentiability of $\pi_1, \ldots, \pi_k$ is assumed in this definition.

Definition 4: The maximum likelihood equation [or m.l.e.] estimate is that root (or a root) of the likelihood equation which provides the maximum of (2.3) when θ is restricted to the roots of (2.4).

Definition 5: A statistic $T(p_1, \ldots, p_k)$ which is a function of relative frequencies only is Fisher consistent [FC] for θ if $T\{\pi_1(\theta), \ldots, \pi_k(\theta)\} = \theta$.

In what follows, a neighbourhood of the point $\pi = (\pi_1, \dots, \pi_k)$ in the space of *all* multinomial distributions is denoted by $N(\pi)$. The notation $p \epsilon N(\pi)$ means that the point $p = (p_1, \dots, p_k)$ is in a neighbourhood of $\pi = (\pi_1, \dots, \pi_k)$.

Lemma 1: *Let θ be the parameter to be estimated. If*

(a) *$T(p_1, \dots, p_k)$ is FC for θ,*

(b) *$T(p_1, \dots, p_k)$ admits continuous first partial derivatives,*

then (i) *$T(p_1, \dots, p_k)$ is asymptotically normally distributed with mean θ and variance*

$$V(T) = \left\{ \sum \pi_i \left(\frac{dT}{d\pi_i} \right)^2 - \left(\sum \pi_i \frac{dT}{d\pi_i} \right)^2 \right\} \div n \qquad \dots \quad (2.5)$$

and (ii)

$$nV(T) \geqslant 1/i(\theta) \qquad \dots \quad (2.6)$$

where $i(\theta) = \sum \frac{1}{\pi_i} \left(\frac{d\pi_i}{d\theta} \right)^2$ *is information as defined by Fisher.*

The conditions (a) and (b) need be satisfied only in the neighbourhood of the true value of θ. For a detailed proof of these results see Rao (1955) or Kallianpur and Rao (1956).

Lemma 2: *For $a_1, \dots, a_k$ positive and $b_1, \dots, b_k$ non-negative such that* $\Sigma a_i = \Sigma b_i$,

$$\sum_1^k a_i \log \frac{b_i}{a_i} \leqslant 0. \qquad \dots \quad (2.7)$$

The equality is attained when and only when $a_i = b_i$ for all i.

We follow a proof similar to that given by Kullback and Liebler (1951) in the case of continuous distributions. The function $\log x$ has the expansion

$$\log x = \log 1 + (x-1) - (x-1)^2 \frac{1}{2y^2}, \; y \epsilon (1, x)$$

On substituting this for each term in the expression (2.7) we have

$$\sum a_i \log \frac{b_i}{a_i} = -\sum a_i \left(\frac{b_i}{a_i} - 1 \right)^2 \frac{1}{2y^2}, \; y_i \epsilon \left(1, \frac{b_i}{a_i} \right) \qquad \dots \quad (2.8)$$

which is always negative and can be zero when and only when $b_i = a_i$ for all i.

3. CONSISTENCY OF M. L. ESTIMATES

3.1. *Estimation of the multinomial distribution under no regularity conditions.* Let A be the admissible set of hypothetical cell frequencies. No assumption is made about this set, which is arbitrary but fixed. The object of estimation is to choose one element of the set as an estimate, given an observed set of relative frequencies $p = (p_1, \ldots, p_k)$. We denote a typical member of A by $\pi = (\pi_1, \ldots, \pi_k)$ and the true value by $\pi^0 = (\pi_1^0, \ldots, \pi_k^0)$. The m.l. estimate is that value of π belonging to A such that $L(\pi)$ is a maximum, with π restricted to A.

Let us suppose first that the m.l. estimate exists at least when $p \epsilon N(\pi^0)$, and represent it by $\hat{\pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_k)$. Then

$$L(\pi^0) \leqslant L(\hat{\pi}) \leqslant L(p) \qquad \ldots \quad (3.1)$$

since $L(p)$ is the unconditional maximum of the likelihood, while $L(\hat{\pi})$ is the maximum when π is restricted to A. By taking logarithms, (3.1) can be written as

$$\Sigma p_i \log \pi_i^0 \leqslant \Sigma p_i \log \hat{\pi}_i \leqslant \Sigma\, p_i \log p_i \qquad \ldots \quad (3.2)$$

It follows from the law of large numbers that (with probability 1) the first and the last sums in (3.2) tend to the common limit $\Sigma\, \pi_i^0 \log \pi_i^0$. Hence, with probability 1

$$\Sigma\, p_i \log \hat{\pi}_i \to \Sigma\, \pi_i^0 \log \pi_i^0 \qquad \ldots \quad (3.3)$$

Without loss of generality we can suppose $\pi_i^0 \neq 0$ for each i, in which case when p is close to π^0, the $\log \hat{\pi}_i$ are bounded, by (3.3). It now follows from (3.3) and the fact that $p \to \pi^0$ that

$$\Sigma\, \pi_i^0 (\log \hat{\pi}_i - \log \pi_i^0) \to 0 \qquad \ldots \quad (3.4)$$

with probability 1. This implies that $\hat{\pi} \to \pi^0$ with probability 1, by an application of (2.8).

Let us now consider the case where the m.l. estimate may not exist, however close p is to π^0. In this case we can prove a more general result that any estimate π^* such that

$$L(\pi^*) \geqslant c \sup_{\pi \epsilon A} L(\pi),$$

where L stands for the likelihood function defined in section 2, and $0 < c \leqslant 1$, is also consistent. This covers the case where a m.l. estimate may not exist for any p but a near m.l. could be chosen. Such approximate m.l. estimates π^* were first considered by Wald (1949).

We have the general relationships

$$L(p) \geqslant L(\pi^*) \geqslant c \sup_{\pi \epsilon A} L(\pi) \geqslant c\ L(\pi^0)$$

Taking logarithms and dividing by n the sample size,

$$\Sigma\ p_i \log p_i \geqslant \Sigma\ p_i \log \pi_i^* \geqslant \frac{\log c}{n} + \Sigma\ p_i \log \pi_i^0$$

As $n \rightarrow \infty$ and $p \rightarrow \pi^0$ we have with probability 1,

$$\Sigma\ p_i \log \pi_i^* \rightarrow \Sigma\ \pi_i^0 \log \pi_i^0$$

and hence by the preceding argument $\pi^* \rightarrow \pi^0$ with probability 1.

The argument employed here can also be used to establish consistency of estimates obtained by other methods of estimation, such as minimum Chi-square, minimum distance etc.

3.2. *Estimation of a parameter.* In general any problem of estimation can be viewed as that of finding the underlying distribution function as a whole, although this is usually done by first estimating a parameter occurring in the distribution function and substituting the estimate for the unknown value. A parameter is a function (or a functional) of the distribution function and for purposes of identification we shall assume that the functional correspondence is one to one. Stated as such the choice of a parameter (i.e. the functional) is to some extent arbitrary. While it was possible to show in section 3.1. that the m.l. estimate of the distribution function is strongly consistent under no regularity conditions it would be of interest to examine whether such a statement of consistency could be made about the m.l. estimate of θ, where $A = \{\pi(\theta)\}$ is a given parametric representation of the set A of admissible distributions.

The answer is in the affirmative provided the parametric representation satisfies the following natural assumption.

Assumption (A_1): The expression

$$I(\theta_0, \theta) = -\Sigma\ \pi_i(\theta_0) \log \frac{\pi_i(\theta)}{\pi_i(\theta_0)}, \quad (I \geqslant 0)$$

which provides an average amount of discrimination between the multinomial distribution defined by θ and the true one defined by θ_0 is bounded away from zero when $|\theta - \theta_0| > \delta$ for each $\delta > 0$.

To prove the assertion made we observe that for any given δ

$$\inf_{|\theta-\theta_0|>\delta} \Sigma\, \pi_i(\theta_0) \log \frac{\pi_i(\theta_0)}{\pi_i(\theta)} = \epsilon > 0$$

Let θ^* be a m.l. or approximate m.l. estimate. Then according to section 3.1. we have (with probability 1)

$$-\Sigma\, \pi_i(\theta_0) \log \frac{\pi_i(\theta^*)}{\pi_i(\theta_0)} < \epsilon$$

for all sufficiently large n. Hence $|\theta^*-\theta_0| \leqslant \delta$ for all sufficiently large n. Since δ is arbitrary, $\theta^* \to \theta_0$ with probability 1.

It is interesting to note that no other regularity conditions are needed. The assumption (A_1) seems to be a natural one to make since its violation would imply that two distributions with close values of the parameter could be better discriminated than those with a larger difference in the parameter which would indeed be an unnatural parametric representation.

Recently Kraft and Lecam (1956) gave an example of a multinomial distribution where the cell frequencies are regular functions of a parameter and for which m.l. estimate of the parameter is not consistent. I am indebted to Dr. R. R. Bahadur who pointed out that the assumption (A_1) introduced in this paper is not satisfied in their case. In fact, in their example, the discrimination between two distributions $\to$ zero as the difference in the corresponding parameter values$\to\infty$.

It may be noted that the corresponding assumption in the minimum Chi-square theory (see Rao, 1955) is that the Chi-square separator

$$\Sigma \frac{[\pi_i(\theta)-\pi_i(\theta_0)]^2}{\pi_i(\theta_0)}$$

be bounded away from zero whenever $|\theta-\theta_0| > \delta$ for any δ.

4. PROPERTIES OF THE M.L.E. ESTIMATE OF THE PARAMETER

We now make the following additional assumptions concerning $\pi_i(\theta)$.

Assumption (A_2): $\pi_1(\theta), \ldots, \pi_k(\theta)$ have continuous partial derivatives of the second order, at least in the neighbourhood of the true value θ_0.

Assumption (A_3): $\pi_j(\theta_0) \neq 0$ for each j, and $(d\pi_j/d\theta) \neq 0$ for at least one j.

As a consequence of this assumption $i(\theta_0)$, which is Fisher's information at θ_0 is $\neq 0$.

Assumption (A_4): $\pi_i(\theta) = \pi_i(\xi)$ for all i implies that $\theta = \xi$.

This provides one to one correspondence between the values of the parameter and the hypothetical cell frequencies

Theorem: *Under the assumptions* $(A_1), (A_2), (A_3)$ *and* (A_4) *there exists a neighbourhood of the true proportions* π^0, *say* $N(\pi^0)$ *and a positive* δ *such that* $p \in N(\pi^0)$ *implies,*

(*i*) *There exists one and only one root* $\hat{\theta}$ *of the likelihood equation* (2.4) *which differs from the true value* θ_0 *by less than* δ. *This root, as a function of the relative frequencies, is continuous at* π^0 *where it tends to* θ_0 *and is Fisher consistent.*

(*ii*) $\hat{\theta}$ *is Frechēt differentiable,*

(*iii*) $\hat{\theta}$ *is the unique m.l. estimate and is therefore in particular the m.l.e. estimate,*

Remark 1: The results (i) and (ii) are true under assumptions (A_2), (A_3) and (A_4) only. In addition the assumption (A_1) is needed to prove the stronger result (iii) that a m.l. estimate exists and is unique at least when p is close to the true value π^0. It may be noted that the assumption A_1 implies A_4.

Remark 2: It follows from the strong law of large numbers that when $\pi = \pi^0$ (i.e., $\theta = \theta_0$)

$$\text{prob. } \{p \in N(\pi^0) \text{ for all sufficiently large } n\} = 1.$$

Consequently the assertions (i), (ii) and (iii) of the Theorem are valid (with probability 1) for all sufficiently large n. A weaker statement but perhaps worthwhile to make is that the probability that the assertions (i)—(iii) are true $\rightarrow 1$ as $n \rightarrow \infty$.

Remark 3: As a consequence of results (i)—(iii) of the Theorem it follows that the m.l. estimate has an asymptotic normal distribution with the least asymptotic variance specified by (2.6) (see Rao, 1955).

Proof of (i) : As $p \rightarrow \pi(\theta_0)$ and $\theta \rightarrow \theta_0$, with $\psi_i = d \log \pi_i/d\theta$, we have

$$\Sigma p_i \frac{d\psi_i}{d\theta} \rightarrow \Sigma\, \pi_i(\theta_0) \left\{ \frac{1}{\pi_i(\theta_0)} \frac{d^2\pi_i}{d\theta_0^2} - \frac{1}{\pi_i^2(\theta_0)} \left(\frac{d\pi_i}{d\theta_0}\right)^2 \right\}$$

$$= -\Sigma \frac{1}{\pi_i(\theta_0)} \left(\frac{d\pi_i}{d\theta_0}\right)^2 = -i(\theta_0) < 0.$$

This implies that there exists a neighbourhood $N(\pi_0)$ for p and positive δ such that $p \in N(\pi_0)$ and $|\theta - \theta_0| < \delta$ implies

$$\Sigma\, p_i \psi_i'(\theta) < 0. \quad \ldots \quad (4.1)$$

More generally we have

$$\Sigma\, p_i \psi_i'(\theta_i) < 0 \quad \ldots \quad (4.2)$$

whenever $p \in N(\pi^0)$ and $|\theta_i-\theta_0| < \delta$. The m.l. equation can be expressed as

$$0 = \Sigma\, p_i\psi_i(\theta_0)+(\theta-\theta_0)\, \Sigma p_i\psi_i'(\theta_i) \qquad \ldots \quad (4.3)$$

where θ_i is in (θ, θ_0). As $p \to \pi^0$, the expression on the right hand side$\to(\theta-\theta_0)\Sigma\pi_i(\theta_0)\psi_i'(\theta_i)$, which is < 0 for $(\theta-\theta_0) > 0$ and > 0 for $(\theta-\theta_0) < 0$ provided $|\theta-\theta_0| < \delta$. Hence when $p \in N(\pi^0)$, $\Sigma\, p_i\psi_i(\theta)$ is positive at $\theta_0-\delta$ and negative at $\theta_0+\delta$. Hence it vanishes for some $\hat\theta$ such that $|\hat\theta-\theta_0| < \delta$.

If there are two roots $\hat\theta_1$ and $\hat\theta_2$ both of which differ from θ_0 by less than δ then it follows, by Rolle's theorem, that

$$\Sigma\, p_i\psi_i'(\hat\theta_3) = 0$$

where $\hat\theta_3$ is in $(\hat\theta_1, \hat\theta_2)$ and hence $|\hat\theta_3-\theta_0| < \delta$. This is impossible when $p \in N(\pi^0)$. Therefore, there is only one root close to θ_0 and the others are separated from θ_0 by a distance greater than δ.

It is now obvious that the root closer to θ_0 is a continuous function of the relative frequencies and $\to \theta_0$ as $p \to \pi^0$. Also it is Fisher consistent at least locally.

Proof of (ii): The expression (4.3) with $\hat\theta$ the continuous root is

$$0 = \Sigma p_i\psi_i(\hat\theta) = \Sigma p_i\psi_i(\theta_0)+(\hat\theta-\theta_0)\Sigma\, p_i\psi_i'(\theta_i)$$

where

$$\theta_i \to \theta_0 \quad \text{as } p_i \to \pi_i(\theta_0).$$

Hence

$$(\hat\theta-\theta_0) = -\Sigma\, p_i\psi(\theta_0)/\Sigma\, p_i\psi_i'(\theta_i)$$

$$= -\Sigma[p_i-\pi_i(\theta_0)]\psi_i(\theta_0)/\Sigma p_i\psi_i'(\theta).$$

Now consider,

$$\frac{1}{\max|p_i-\pi_i(\theta_0)|}\left|(\hat\theta-\theta_0)+\frac{1}{i(\theta_0)}\sum \frac{p_i-\pi_i(\theta_0)}{\pi_i(\theta_0)}\frac{d\pi_i}{d\theta_0}\right|$$

$$= \sum \frac{|p_i-\pi_i(\theta_0)|}{\max|p_i-\pi_i(\theta_0)|}\left|-\frac{\psi_i(\theta_0)}{\Sigma p_i\psi_i'(\theta_i)}+\frac{1}{i(\theta_0)}\frac{1}{\pi_i(\theta_0)}\frac{d\pi_i}{d\theta_0}\right|$$

$$\leqslant \sum\left|\frac{\psi_i(\theta_0)}{\Sigma p_i\psi_i'(\theta_i)}-\frac{1}{i(\theta_0)}\frac{1}{\pi_i(\theta_0)}\frac{d\pi_i}{d\theta_0}\right| \to 0$$

as max $|p_i-\pi_i(\theta_0)|\rightarrow 0$. The function $(\hat{\theta}-\theta_0)$ is thus Frechet differentiable and can be approximated by the linear functional

$$-\frac{1}{i(\theta_0)}\sum\frac{p_i-\pi_i(\theta_0)}{\pi_i(\theta_0)}\frac{d\pi_i}{d\theta_0}$$

and hence $\sqrt{n}(\hat{\theta}-\theta_0)$ has asymptotic normal distribution with zero mean and variance $1/i(\theta_0)$ which is the minimum attainable under the conditions of lemma 1.

Proof of (iii): The proof consists of two parts

(A) For any θ such that $|\theta-\theta_0|<\delta$ and $p\in N(\pi^\circ)$, in which case $|\hat{\theta}-\theta_0|<\delta$,

$$\Sigma\, p_i \log \pi_i(\theta)=\Sigma\, p_i\log\pi_i(\hat{\theta})+\frac{(\hat{\theta}-\theta)^2}{2}\,\Sigma p_i\,\psi_i(\theta')\leqslant\Sigma p_i\log\pi_i(\hat{\theta})\ \ldots\quad(4.4)$$

since by (4.1), $\Sigma\, p_i\,\psi_i'(\theta')$ is negative.

(B) For θ outside $|\theta-\theta_0|<\delta$, let S be the set of θ such that

$\Sigma(\pi_i(\theta_0)-\epsilon')\log\pi_i(\theta)<\Sigma(\pi_i(\theta_0)+\epsilon')\log\pi_i(\theta_0)$, where ϵ' is a small positive number such that the cube $\pi_i(\theta_0)\pm\epsilon'$ is within $N(\pi^0)$. Then, for θ in S and p in the cube,

$$\begin{aligned}\Sigma p_i\log\pi_i(\theta)&\leqslant\Sigma(\pi_i(\theta_0)-\epsilon')\log\pi_i(\theta)\\&<\Sigma(\pi_i(\theta_0)+\epsilon')\log\pi_i(\theta_0)\leqslant\Sigma p_i\log\pi_i(\theta_0).\end{aligned}$$

Now consider θ outside $|\theta-\theta_0|<\delta$ and S. If $\epsilon'<\min\{\pi_i(\theta_0)\}$ the quantities $\log\pi_i(\theta)$ are bounded. Also

$$\Sigma\, p_i\log\frac{\pi_i(\theta)}{\pi_i(\theta_0)}=\Sigma\pi_i(\theta_0)\log\frac{\pi_i(\theta)}{\pi_i(\theta_0)}+\Sigma[p_i-\pi_i(\theta_0)]\log\frac{\pi_i(\theta)}{\pi_i(\theta_0)}\ \ldots\quad(4.5)$$

$$\leqslant -c+\nu\qquad\ldots\quad(4.6)$$

where by assumption (A_1) the first expression on the right side of (4.5) is not greater than say, $-c<0$, and $\nu\rightarrow 0$ as $p\rightarrow\pi^0$. Hence (4.6) can be made <0 by choosing $p\in N(\pi^\circ)$. Thus when $|\theta-\theta_0|\geqslant\delta$

$$\Sigma\, p_i\log\pi_i(\theta)<\Sigma\, p_i\log\pi_i(\theta_0)\leqslant\Sigma\, p_i\log\pi_i(\hat{\theta})\qquad\ldots\quad(4.7)$$

the latter inequality being true in virtue of what is proved in (A).

Since the equality in (4.4) is attained when and only when $\theta=\hat{\theta}$ for $|\theta-\theta_0|<\delta$ and outside $|\theta-\theta_0|<\delta$ the inequality (4.7) is strictly true, it follows that $\hat{\theta}$ is the unique m.l. estimate and in particular it is the unique m.l.e. estimate,

In conclusion, I wish to thank my colleague Dr. R. R. Bahadur for the useful discussions I had with him during the preparation of this paper.

REFERENCES

CRAMÉR, H. (1946): *Mathematical methods of statistics*, Princeton University Press.

HUZURBAZAR, V. S. (1948): The likelihood equation, consistency and the maximum of the likelihood function. *Ann. Eugen.*, **14**, 185.

KALLIANPUR, G. and RAO, C. R. (1955): On Fisher's lower bound to the asymptotic variance of an estimate. *Sankhya*, **15**, 331.

KRAFT, C. AND LECAM, L. (1956): A remark on the roots of the maximum likelihood equation. *Ann. Math. Stat.*, **27**, 1174.

KULLBACK, S. AND LEIBLER, R. A. (1951): On information and sufficiency. *Ann. Math. Stat.*, **22**, 79.

RAO, C. R. (1955): The theory of the method of minimum chi-square estimation. *Proc. Int. Stat. Conf.* 1955 (in press).

WALD, A. (1949): Note on the consistency of the maximum likelihood estimation. *Ann. Math. Stat.*, **20**, 595.

Paper received : May, 1957.

MISCELLANEOUS

MODERN TRENDS IN TIME SERIES ANALYSIS

By ULF GRENANDER
University of Stockholm

INTRODUCTION

For a long time there has been considerable dissatisfaction with the methods designed for the statistical analysis of time series. The methods are said to be irrelevant to most of the situations met with in practice; one does not have precise conditions under which they are valid; the underlying mathematical proofs are incomplete or incorrect; the numerical work involved is too cumbersome or the efficiency, power etc., is too small. Many time series analysts will agree that this criticism is at least partly true.

During the last decades the theory has expanded at a fast rate, and more important the general outlook has changed basically due to the impact from various applied fields and because of the advances of the modern theory of probability. Therefore it is tempting to sum up the present situation in this field, to evaluate the techniques we have at our disposal and to find out if the above criticism really applies and if so what can be done to remedy the defects of the theory. As far as possible I will concentrate on the main ideas and try to avoid going into technical details.

1. PROBABILITY MODELS

Suppose we observe a quantity x which varies with time. Either we measure x continuously for all values of the time parameter t in some interval or only at certain discrete true points, say the first day of very month. We can get continuous data, yearly data, monthly data and so on. The series of values obtained will be denoted $x(t)$ and the set of the time points at which the measurements are made is called T. We say that $x(t)$ is a *time series* measured on T.

In statistics we deal with measurements of stochastic variables and want to use these measurements or samples to make statements about the underlying probability distributions. In the present case we thus consider $x(t)$ as values obtained from a collection of stochastic variables $X(t)$ where t runs through T.

For each fixed t we have a stochastic variable $X(t)$ which can be described by its distribution function. To characterize the whole collection of stochastic variables we must also consider the simultaneous distributions.

We have to describe the probability distributions of $X(t)$ and this problem belongs, as is well known, to the theory of *stochastic processes*. If T is *a whole* interval we deal with stochastic processes with continuous time parameter, if T is a finite or denumerable set of values we consider stochastic processes with discrete time parameter.

To determine what classes of probability distributions we should use to form the stochastic processes we must be guided by the demands from the applied fields. In a specific case we may have enough knowledge of the random mechanism that generates $X(t)$ to describe the stochastic process more or less completely. Sometimes we can state that the true probability distribution belongs to a fairly small class characterized by a few parameters. But often the natural class is much wider and involves a large or even infinite number of unknown parameters. From a certain point of view the latter case is typical for the problems one meets to-day in time series analysis.

It will be convenient to introduce the following terminology. Denote the expected value of the stochastic variable $X(t)$ by $m(t)$,

$$m(t) = EX(t).$$

We then write

$$X(t) = m(t)+\gamma(t)$$

and we may call $m(t)$ the *systematic component* and $\gamma(t)$ the *random component* of the stochastic process studied. It is clear that we cannot decide which method of analysis is appropriate before we have chosen the class of probability distributions which is thought to correspond to the practical problem. As we shall see the probability models that have been proposed fall into three categories that roughly correspond to three stages of the historical development of time series analysis. I admit that these stages are loosely defined and overlap both in time and logical content.

2. INDEPENDENT RANDOM COMPONENTS

The earlier attacks on these problems were usually based upon the assumption, made explicitly or not, that the random components $\gamma(t)$ were *independent* (and may be identically distributed) stochastic variables. The systematic component $m(t)$ was often given as a linear expression of functions known except for some parameters. A typical example and at the same time one of the most important of these models is the following :

$$m(t) = \sum_{\nu=1}^{r} A_\nu \cos\left(\frac{f_\nu t}{2\pi}+\phi_\nu\right) \quad (t = 1, 2, \dots n)$$

where the amplitudes A_ν, the frequencies f_ν and the phases ϕ_ν are fixed numbers some of which may be unknown and the rest given a priori. If the random components $\gamma(t)$ are supposed to be independent and normally distributed with mean value zero and a common variance σ^2, methods have been deviced to estimate the parameters and to test various hypotheses. The statistic often used for this purpose is the so called *periodogram*

$$I(f) = \frac{1}{2\pi n}\left[\sum_{t=1}^{n} x_t \cos\frac{ft}{2\pi}\right]^2+\frac{1}{2\pi n}\left[\sum_{t=1}^{n} x_t \sin\frac{ft}{2\pi}\right]^2.$$

Fisher's periodogram test (Fisher, 1929) is the most widely used of these methods. In general these techniques aim at detecting and verifying the true periods $1/f_\nu$ and sometimes one wishes to estimate the associated amplitudes A_ν. Earlier scientists had used the periodogram in an uncritical way in their enthusiastic search for periods everywhere. When Fisher's test and related statistical tools became available a more careful analysis of data of this kind was made possible.

Many other methods of this type have been invented, some refining and generalizing the estimates and tests based on the periodogram and also some dealing with other types of systematic components. Among the latter I will just mention those based on the assumption that $m(t)$ is a polynomial with unknown coefficients and perhaps of unknown degree.

It is evident that the problems sketched in this section belong to the theory of linear hypotheses embodying analysis of variance, regression analysis etc.

The models with independent random components have been used in many fields with varying success. It has been recognized for a long time that the basic assumption of independence breaks down often when the time between successive measurements is short compared to the time scale of the random mechanism. If the dependence is not negligible it can be dangerous and quite misleading to use methods valid only for the models assuming independence. It may be instructive to consider the following almost trivial example.

$$\gamma(t) = \text{normal with mean zero and known variance } \sigma^2.$$

$$m(t) = \text{an unknown constant } m.$$

We want to estimate m using a sample $x_1, x_2, \dots x_n$. The usual estimate

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x(t)$$

is then always unbiased. If the γ's are independent we know that $\bar{x}$ has the standard deviation $\sigma/\sqrt{n}$. Hence we can form a confidence interval for m, and this interval will have a length of the order of magnitude $\sigma/\sqrt{n}$. On the contrary when there is dependence, this is not necessarily true. E.g. in the extreme case of complete dependence $\gamma(1) = \gamma(2) = \dots = \gamma(n)$ we get clearly a standard deviation σ for $\bar{x}$. Hence, if we use the above confidence interval in the latter case, we will make a much more precise statement than what we are really entitled to do.

Things are perhaps not quite so extreme in practice but it is clear that a careful examination is necessary before one should apply these methods to situations, where the independence is not *a priori* clear.

3. Finite parameter schemes

To handle such difficulties other probability models have been designed to obtain methods of analysis based on more valid assumptions. The pioneer work in this direction is that of U. Yule (1927) and H. Wold (1938). These authors consider models, where the random component $\gamma(t)$ is a so-called *finite parameter scheme*. It will be sufficient for our purpose to mention one of these, the autoregressive process.

$\gamma(t)$ is such a process if it satisfies a linear difference equation with constant coefficients

$$a_0\gamma(t)+a_1\gamma(t-1)+a_2\gamma(t-2)+\dots+a_l\gamma(t-l) = \xi(t)$$

where l is called the order of the scheme. Here the stochastic variables $\xi(t)$ are independent and identically distributed and are usually normed to have mean value zero. It is easy to verify that the γ's are not independent now. In fact, introducing the *covariance function*

$$r(s, t) = E\gamma(s)\,\gamma(t),$$

which can be computed without difficulty, one finds that $r(s, t)$ does not vanish generally when $s \neq t$ as it would if the variables were independent. In applied work l is usually chosen as 1 or 2.

In certain applications one is naturally led to stochastic processes satisfying difference equations of this type. This is somewhat analogous to the traditional problems in physics which are often reduced to second order differential equations.

Of course it must be very important to be able to distinguish empirically between stochastic processes which have independent random components and those of the present type. For this purpose a collection of methods were designed mainly in the U.S. during the 1940's beginning with the papers of T. Koopmans (1942) and R. L. Anderson (1942). Techniques of a somewhat different kind were developed by Quenouille and H. Wold (1949) to test whether a given time series belonged to one of the finite parameter schemes. Finally I should mention a great deal of results on the sampling distribution of various statistics to be used in relation to these schemes (M. S. Bartlett, 1946). Much work is going on in this direction and important investigations with bearings on the finite parameter schemes have been carried out the last years.

In spite of these achievements some criticism has been directed against this approach. It is not clear, it is argued, that the finite parameter schemes constitute the correct and natural class of processes for the applications one is interested in. Although it is true in a certain sense that a very wide and important class of processes can be approximated as nearly as one might wish by finite parameter schemes of high order, one does seldom have any reason to believe that a scheme of the first or second order is enough to get a reasonable approximation. Why not start instead from this wider class ? The supporters of this idea point out that by restricting one self *a priori* to these low order schemes, the statistician places himself in a situation similar to that discussed before : he is making statements more precise than those he is really entitled to make. In other words by using information that he really does not have, he gets an illusory feeling of security when he draws his statistical conclusions.

4. MORE GENERAL MODELS

In the last decade much work has been devoted to the invention of methods of more general scope. I will speak only of one approach, viz. the one via the stationary stochastic processes, as the stationary case is the one studied most extensively so far. Loosely speaking a process is said to be *stationary* when its probability distribution does not change with time, so that its probability distributions do not depend upon absolute time t, only upon time differences. Hence $\gamma(t)$ is a stochastic variable with a distribution function independent of t. The simultaneous distribution function of the two stochastic variables $\gamma(s)$ and $\gamma(t)$ depend only upon the time difference $t-s$ but not upon t or s separately, and so on for several variables. One can expect to find a stationary model suitable when one knows that the random mechanism behaves in a certain invariant way; this has to be investigated from case to case. It is clear that even a stable random mechanism can generate a process of transient nature although in many cases a *probabilistic equilibrium* is reached after a sufficiently long time.

Although the stationarity assumption restricts the applicability of the model to a great extent, the class of its possible probability distributions is still very large, too large for many purposes. Various subclasses have been studied, and for convenience I will only discuss those stationary processes which are normally distributed, but much of what I say is applicable much more generally.

As is well known a normal probability distribution is completely characterized by its moments of the first and second order. $\gamma(t)$ has mean value zero and its covariance function $r(s, t)$ depends only upon $s-t$ so that we can write

$$r(s, t) = R(s-t).$$

In other words if we know $R(t)$ for all values of t then we know the probability distribution completely. From a sample we can estimate $R(t)$ and construct confidence regions, at least when the sample is large.

While the approach via the finite parameter schemes has been influenced mainly but not exclusively by applications to economics and the social sciences, the methods I now speak of have been developed mainly to solve statistical problems in the physical sciences. In many of these contexts it turned out that the quantity of immediate interest was not $R(t)$ but a related function, the so called spectral distribution function F. According to a famous theorem of Wiener-Khinchin (see e.g. J. L. Doob, 1953) there is a function $F(\lambda)$ with all the properties of a distribution function except that $F(\infty)$ is not 1 but equal to $R(0) = EY^2(t)$ and such that

$$R(t) = \int_0^\infty \cos t\lambda dF(\lambda).$$

We are now speaking of the case of a continuous parameter when t runs through all real values, but in the discrete case one has instead according to H. Wold (1938)

$$R(t) = \int_0^\pi \cos t\lambda dF(\lambda).$$

Hence if the function $F(\lambda)$ is known we can calculate $R(t)$, and the converse is also possible through the use of an inversion formula.

The role of the spectral distribution function is made clearer by an important theorem of H. Cramér (1942) which asserts that any stationary process $y(t)$ can be represented by an expression

$$Y(t) = \int_0^\infty \sin t\lambda dZ_1(\lambda) + \int_0^\infty \cos t\lambda dZ_2(\lambda)$$

where $Z_1(\lambda)$ and $Z_2(\lambda)$ are stochastic processes with the properties

(i) $EZ_1(\lambda) = EZ_2(\lambda) = 0$.

(ii) $EZ_1(\lambda)Z_2(\mu) = 0$, the two processes are uncorrelated.

(iii) $E[Z_1(\lambda_2)-Z_1(\lambda_1))][Z_1(\lambda_4-\lambda_3)] = 0$ if the two intervals (λ_1, λ_2) and (λ_3, λ_4) are disjoint; the analogous relation for $Z_2(\lambda)$.

(iv) $E[Z_1(\lambda_2)-Z(\lambda_2))]^2 = E[Z(\lambda_2)-Z(\lambda_1)]^2 = F(\lambda_2)-F(\lambda_1)$ for any $\lambda_1 < \lambda_2$

This representation reminds one both of ordinary Fourier integrals and of the trigonometric expressions discussed in section 2. There is an essential distinction however, but to avoid technicalities we will not go into a detailed discussion of how the above integrals should be interpreted. Instead we remark only that, heuristically speaking, the process is decomposed into harmonic terms sin $t\lambda$ and cos $t\lambda$ with stochastic amplitudes $\Delta Z_1(\lambda)$ and $\Delta Z_2(\lambda)$ respectively. The order of magnitude of these amplitudes in given by their variances

$$\text{var}\ (\Delta Z_1) = \text{var}\ (\Delta Z_2) = \Delta F.$$

In many context a variance can be interpreted as average power; a quantity of immediate physical meaning which can be measured directly.

It is then natural to try to estimate $F(\lambda)$ directly, or perhaps its derivative $f(\lambda) = \frac{d}{d\lambda} F(\lambda)$. Various methods were tentatively suggested by J. W. Tukey, M. S. Bartlett (1950, 1955) , and U. Grenander (1951) and much work is going on in this direction. It is interesting to note that several of the important recent results in this branch of theoretical statistics are due to research workers in the physical sciences. This has influenced the development of this theory and increased the practical usefulness of the suggested methods. It deserves to be mentioned that the periodogram in modified form reappears in this context.

In connection with the estimation of an unknown function such as $F(\lambda)$ one asks naturally if it is possible to construct confidence regions of the type

$$F_1{}^*(\lambda) < F(\lambda) < F_2(\lambda) \text{ for all } \lambda.$$

This and related problems was treated by U. Grenander—M. Rosenblatt (1953) where large sample results were obtained.

A group of problems of a somewhat different kind arise when we want to estimate unknown quantities determining the systematic component $m(t)$ when $y(t)$ is a stationary process. The classical results of regression analysis break down, but fairly definite results can be obtained (see. U. Grenander 1954); these are however also of a large sample character, and the degree of approximation for small samples has not been thoroughly studied so far.

Among the fields where the above methods have been applied we may mention the theory of noise and the study of hydrodynamic turbulence. Sometimes, and especially when analog devices are used, it is convenient to use models with a continuous time parameter. One is then confronted by certain difficulties but it was shown by U. Grenander (1950) that they can be overcome by using the modern theory of stochastic processes as formulated by J. L. Doob. Work in this direction has been done by H. B. Mann (1953) and others.

It is impossible to discuss or even mention all the recent contributions. Many of them are published in non-statistical journals and are not always easy to locate. In a forthcoming monograph Dr. M. Rosenblatt and I will present these methods in a more systematic manner and discuss their applicability.

5. The present situation

Even this short resume should be enough to show what a wealth of methods that are now available in time series analysis. To find out the relation between them we shall examine the three main groups of models.

The first, with independent random components, has been severely criticized in the literature. It is clear though, that this criticism is directed more against certain uncritical applications of them than against the methods *per se* : they are correct and sound under the appropriate conditions. After all, these techniques and models belong to the traditional set up in theoretical statistics.

With respect to the second group, the finite parameter schemes, it is the same thing; the methods are valid in a certain domain and the question is only if our information collected earlier empirically or deductively enables us to state that the probability structure is of this particular type.

The third group finally contains methods of great generality and applicability, this should not lead us to forget that they too are designed for certain wide but limited classes of stochastic processes. The indiscriminate use of these techniques will be as dangerous as to apply one of the earlier methods outside their domain of validity.

Take e.g. the stationarity assumption, (of the random component) which is often made and is so temptingly convenient for the theoretician. While it is fairly well established in certain problems arising in the physical sciences, one suspects that e.g. economic time series seldom are generated by a stationary mechanism. At least this cannot hold over long time intervals.

And how about the assumption of normality or other similar but more general conditions. As soon as we do not have *a priori* information on this question, we are in a difficult situation since empirical verification presents a complicated theoretical problem and at any rate demands a lot of observations. What we need is possibly a test or, more likely, a way of estimating the probability distributions with confidence regions.

An unpleasant aspect common to most of the recent developments is that a large amount of numerical work is needed to compute the statistics. When large scale or routine investigations are made of time series, this may not be so serious, since electronic computers and punched card calculators in current use have proved well adapted for this purpose. It is much worse when just a single or a few time series should be analyzed, so that it is unpractical to devote time to the coding of the numerical problem. Then one would need some convenient short cut method, but at present we know very little of such possibilities.

Much of the recent work in time series analysis has been done to invent rather general tools based on realistic assumption. The gain we get in extending the validity must be compensated by a loss of power or efficiency. Or more precisely: if one of these methods is applied to a situation where the possible probability distributions form a more restricted class, then we can expect that we will not be able to make as precise statements as if we had applied a technique specially designed for this latter class of models. If we insist in demanding moderate or high precision in our estimates, quite long time series will be required or alternatively a fair number of shorter series. In connection with this it may be pointed out that most of the recent work is of a large sample character.

As we have seen the three groups of models complement each other and correspond to different practical problems. Although much progress has been made in time series analysis, it is evident that much has to be done before the theory can be said to be in a satisfactory shape.

The present fairly general approach to time series analysis is somewhat hampered by the lack of suitable families of stochastic processes defined to fit the demand from the applications. Therefore the most important need is perhaps that of more realistic and flexible probability models. To do this we have to rely on the applications : close contact with the subject sciences will help us to construct such classes of stochastic processes that are well fitted for practical use. The success of some of the new techniques mentioned in Section 4 is undoubtedly due to the fact that they are founded on models with a sound deductive background. Instead of just trying to fit a model mechanically to the data, one should seek one based on a more penetrating analysis. The class of probability structures ought to correspond to the knowledge we already have of the phenomena under investigation. Causal concepts may be helpful for this construction. An illuminating discussion of this and related topics can be found in H. Wold (1952).

It seems likely that also in future work in time series analysis we will have to be satisfied with approximate methods. I would be surprised if it showed possible to obtain methods of inference valid exactly for small samples under such general conditions that we are interested in. In the light of what has been said earlier, this need not depress any except the perfectionists, since we will need fairly large samples anyway to get sufficient resolving power in the analysis. The modern tendency among theoretical statisticians of to-day towards abstract methods, valuable and necessary also in time series analysis, has resulted in that problems with a more practical flavour have been neglected to some extent. The extensive formal manipulations may scare the theoretician and offend his aesthetic sense, but there are a lot of interesting, difficult and tempting questions involved, usually of an analytic character, also when approximate methods should be designed. This is so at least if one has ambitions to get logically sound and mathematically proved results.

This brings me to another point which I consider important. It has been argued that, as statistics is an entirely applied science, it is not necessary to have completely rigorous derivations. Now the statistician usually takes the attitude that one of this jobs (and perhaps the main one) is to spread the use of sound techniques in research when statistics is used in practice. We all know how easy it is to make an error in the planning and analysis of statistical investigations. The statistician should see that this does not happen, and he is then helped by his experience and by statistical theory. If this should work it is necessary that the theory is correct. Now, in many branches of statistics the mathematical difficulties are negligible compared to the questions of application and interpretation of the method. If the algebra and calculus involved is not used in a quite rigorous way, there still is not very great danger of committing a grave error, if the method is based upon a sound intuitive argument. However as the mathematical complexity increases, as has happened in time series analysis and which is likely to go on, then one has to be more careful with the mathematics. There are examples in time series analysis where errors of this kind have led to misleading or erroneous results. I do not want to underestimate the role of intuition in scientific work, but we should be careful before offering a new statistical technique, so that we know if and when it works. It is only a simple act of intellectual honesty to apply the same norms on the logical basis for theoretical work as when we criticize the subject scientist for his defect statistical reasonings. We have all met the holier-than-thou attitude among some statisticians towards their colleagues in applied fields. Furthermore a correct mathematical proof interpreted in the right manner tells us a good deal about the range of the methods: proofs are not a luxuary but can and should be used to tell us practical things. Also let us note that the practical statistician does not have the same possibilities of empirically verifying or falsifying the validity of his method as does *e.g.* the physicist. The only way in which the statistician can do this is by artificial sampling experiments. This is quite costly and cumbersome to do in time series analysis, at least that is my experience.

I would like to discuss briefly another topic of considerable importance, how to use time series for the purpose of prediction ? The theory of prediction due to Wiener-Kolmogorov (see Wiener 1949) is well-known. It is designed for the case when we have observed a stationary process (say the one we have denoted by $y(t)$) up till some time t. We want to predict the value of the process at some future time point, $y(t+h)$. This beautiful theory has many important applications, but in many practical problems it cannot be used since either $y(t)$ is not stationary or because there is an unknown systematic component $m(t)$ superimposed on the observations. If we do not have a good idea of the underlying mechanism, we may be unable to describe $m(t)$ closely enough. This is certainly a case where overenthusiastic use of existing tools of time series analysis might lead to disastrous results.

Situations like this exist for long range forecasting in fields like economics, demography and others. It is clear that the solution does not lie only in a refined statistical treatment but more information on the functioning of the generating mechanism must be collected so that the structure of $m(t)$ and $y(t)$ can be described in more precise terms. Statistics does not work miracles just because there is a superstitious belief in it.

6. Conclusions

Summing up I would put the emphasis on the following items :

(i) Keeping in touch with the practical demands the time series analysts and the probabilists together must invent new classes of stochastic models, based on more realistic assumptions about the underlying mechanism.

(ii) Starting from the appropriate models statistical methods should be developped, which are applicable for at least large and moderate sample size. The validity of the approximations should be examined. Attention should be given to the need for simpler methods.

(iii) The theoretical investigations should be based on a rigorous logical and mathematical foundation.

REFERENCES

HARTLEY, H. O. (1943-46): Studentization, or the elimination of the standard deviation of the parent population from the random sample distribution of statistics. *Biometrika*, **33**, 173-180

MORIGUCHI, S. (1953): A note on Hartley's formula of studentization. *Rep. Stat. App. Res.*, **2**, 1-5.

NAIR, K. R. (1948): The distribution of the extreme deviate from the sample mean and its studentized from. *Biometrika*, **35**, 118-144.

——— (1948): The studentized form of the extreme mean square test in the analysis of variance. *Biometrika*, **35**, 16-31.

——— (1952): Tables of percentage points of the "Studentized" extreme deviate from the sample mean. *Biometrika*, **39**, 189-191.

NATIONAL BUREAU OF STANDARDS (1951): *Tables of* $n!$ *and* $\Gamma(n+\frac{1}{2})$ *for the First Thousand Values of* n.

PEARSON, E. S. and HARTLEY, H. O. (1943-46): The tables of the probability integral of the studentized range. *Biometrika*, **33**, 89-99.

Paper received : December, 1955.

A CLASS OF TWO-DIMENSIONAL RANDOM VARIABLES AND DISTRIBUTION FUNCTIONS

By H. S. KONIJN[1]

University of California, Berkeley

SUMMARY

This paper considers some properties of the class of two-dimensional random variables which have been obtained by rotation of two independent random variables and of the corresponding class of distribution functions. Normal distributions play a distinguished role in this class of distributions. When second moments are finite, the case of equal variances and the two cases of the first variate having a bigger variance than the other or a smaller one are considerably different. Thus, in the first case the correlation among the rotated variables is zero, whether the variables are dependent or not (excepting a rotation by ±45°); in the latter two cases the correlation vanishes only when the variables are independent and the correlation coefficient and the rotation angle are monotone functions of each other, the sense depending on the sign of the difference between the variances. Provided the angle rotation is and remains different from ±45°, further rotation cannot change the threefold classification here given. It is also natural to examine the consequences of and the conditions for symmetry of the distributions of the variables involved.

1. INTRODUCTION

For Y_0 and Z_0 independent random variables with joint distribution F_0 and marginals G and H, and (to avoid trivialities) $|\theta| \leqslant \pi/4$, let Y_θ and Z_θ denote

$$Y_\theta = Y_0 \cos\theta + Z_0 \sin\theta,$$

$$Z_\theta = -Y_0 \sin\theta + Z_0 \cos\theta,$$

F_θ the joint distribution of Y_θ and Z_θ, and $R\,[F_0]$ the class of distributions thus obtained. In the case of one-sided alternatives to F_0, one can consider the class obtainable by rotations in a direction increasing [decreasing] the correlation, provided the correlation coefficient exists and is not identically zero; then one may, if desired, distinguish notationally $R^+[F_0]$ and $R^-[F_0]$. (In contrast to the angle of rotation, the correlation coefficient does not depend on the scale parameters of the marginal distributions.)

We here consider properties of $R\,[F_0]$ for their intrinsic interest; the local asymptotic power of various tests for independence against alternatives in some union of classes $R\,[F_0]$ has been examined elsewhere (Konijn).

---

[1] At Present in the University of Sydney, Australia.

2. ROLE OF THE NORMAL DISTRIBUTION

Its role is brought out by Theorem 2.1.

Theorem 2.1: *For $\theta \neq 0$, F_θ is either normal with $\rho = 0$ and $\sigma^2(Y_\theta) = \sigma^2(Z_\theta)$, or Y_θ and Z_θ are dependent random variables.*

This is an immediate application of Theorem 2.1 of (Konijn).

3. CONSEQUENCES OF EQUAL OR UNEQUAL VARIANCES

The next two propositions differentiate the cases of equal and unequal variances of Y_θ and Z_θ.

Theorem 3.1: *Let $0 < \sigma(Y_\theta) = \sigma(Z_\theta) = \sigma < \infty$ for some θ. If either*

(a) $|\theta| \neq \pi/4$, *or*

(b) $\rho(Y_\theta, Z_\theta) = 0$, *then $\sigma(Y) = \sigma(Z) = \sigma$ and $\rho(Y, Z) = 0$ for all Y and Z having a joint distribution $F \epsilon R[F_0]$; while if*

(c) $\rho(Y_\theta, Z_\theta) \neq 0$,

then
$$|\theta| = \pi/4 \text{ and } \sigma(Y_0) \neq \sigma(Z_0).$$

Proof: If $\rho(Y_\theta, Z_\theta) = 0$,

$$\sigma^2(Y_0) = \sigma^2(Y_\theta)\cos^2\theta + \sigma^2(Z_\theta)\sin^2\theta - \sigma(Y_\theta)\,\sigma(Z_\theta)\,\rho(Y_\theta, Z_\theta)\sin 2\theta,$$

which equals σ^2, and similarly $\sigma^2(Z_0) = \sigma^2$.

So for $F_{\theta'} \epsilon R[F_0]$,
$$\sigma^2(Y_{\theta'}) = \sigma^2(Y_0)\cos^2\theta' + \sigma^2(Z_0)\sin^2\theta' = \sigma^2,$$

and
$$\sigma^2(Z_{\theta'}) = \sigma^2,$$

and
$$\rho(Y_{\theta'}, Z_{\theta'}) = (2\sigma^2)^{-1}\{\sigma^2(Z_0) - \sigma^2(Y_0)\}\sin 2\theta' = 0, \text{ proving (b).}$$

Now
$$0 = \text{cov}(Y_0, Z_0) = \sigma^2\rho(Y_\theta, Z_\theta)\cos 2\theta + \tfrac{1}{2}\{\sigma^2(Y_\theta) - \sigma^2(Z_\theta)\}\sin 2\theta$$
$$= \sigma^2\rho(Y_\theta, Z_\theta)\cos 2\theta$$

implies that $\rho(Y_\theta, Z_\theta) = 0$, when $|\theta| \neq \pi/4$, so that also (a) and the first part of (c) are proved.

If
$$\theta = \pi/4,$$
$$\sigma^2(Y_0) = \sigma^2 - \sigma^2\rho(Y_\theta, Z_\theta)\sin 2\theta = \sigma^2\{1 - \rho(Y_\theta, Z_\theta)\},$$
$$\sigma^2(Z_0) = \sigma^2 + \sigma^2\rho(Y_\theta, Z_\theta)\sin 2\theta = \sigma^2\{1 + \rho(Y_\theta, Z_\theta)\};$$

if
$$\theta = -\pi/4,$$
$$\sigma^2(Y_0) = \sigma^2\{1 + \rho(Y_\theta, Z_\theta)\},$$
$$\sigma^2(Z_0) = \sigma^2\{1 - \rho(Y_\theta, Z_\theta)\}.$$

Theorem 3.2: *Let $0 < \sigma(Y_\theta) < \sigma(Z_\theta) < \infty\,[\sigma(Y_\theta) > \sigma(Z_\theta)]$ for some θ.*

(a) $|\theta| \neq \pi/4$, *and* $\tan 2\theta = 2\rho(Y_\theta, Z_\theta)\,\sigma(Y_\theta)\,\sigma(Z_\theta)/\{\sigma^2(Z_\theta) - \sigma^2(Y_\theta)\}$;

and $\rho(Y_\theta, Z_\theta) = 0$ *implies* $\theta = 0$.

(b) $\sigma(Y) < \sigma(Z)[\sigma(Y) > \sigma(Z)]$ *for all* Y *and* Z *having a joint distribution* $F \in R[F_0]$ *different from an* F^c, *and* $\sigma(Y) = \sigma(Z)$ *for* Y *and* Z *having for joint distribution an* F^c. *(Here* F^c *denotes the elements of* $R[F_0]$ *for which* $|\theta| = \pi/4$.*)*

(c) $\rho(Y_{\theta'}, Z_{\theta'}) = \rho'$ *and* θ' *are increasing [decreasing] continuously differentiable functions of each other for*

$$0 \leqslant \rho' \leqslant \{\sigma^2(Z_0) - \sigma^2(Y_0)\}/\{\sigma^2(Z_0) + \sigma^2(Y_0)\}$$

$$= \frac{\{\sigma^2(Z_\theta) - \sigma^2(Y_\theta)\}\cos 2\theta + 2\rho(Y_\theta, Z_\theta)\,\sigma(Y_\theta)\,\sigma(Z_\theta)\sin 2\theta}{\sigma^2(Z_\theta) + \sigma^2(Y_\theta)} \text{ and } 0 \leqslant \theta' \leqslant \pi/4$$

$$[0 \leqslant \rho' \leqslant -\{\sigma^2(Z_0) - \sigma^2(Y_0)\}/\{\sigma^2(Z_0) + \sigma^2(Y_0)\} \text{ and } 0 \geqslant \theta' \geqslant -\pi/4]$$

in $R^+[F_0]$, *and for the mirrored ranges in* $R^-[F_0]$. *The functions vanish at* $\theta' = 0$ *and* $\rho' = 0$.

Proof: $|\theta| \neq \pi/4$, since otherwise we would have

$$\sigma^2(Y_\theta) = \tfrac{1}{2}\{\sigma^2(Y_0) + \sigma^2(Z_0)\} = \sigma^2(Z_\theta)$$

contrary to hypothesis.

The rest of (a) follows from the relations

$$0 = \operatorname{cov}(Y_0, Z_0) = \operatorname{cov}(Y_\theta, Z_\theta)\cos 2\theta - \tfrac{1}{2}\{\sigma^2(Z_\theta) - \sigma^2(Y_\theta)\}\sin 2\theta.$$

To prove (b), let $\sigma(Y_\theta) = \sigma$, $\sigma(Z_\theta) = \alpha\sigma$, $\rho(Y_\theta, Z_\theta) = \rho$.

Then

$$\sigma^2(Y_0) = \sigma^2\{\cos^2\theta + \alpha^2\sin^2\theta - \alpha\rho\sin 2\theta\},$$

$$\sigma^2(Z_0) = \sigma^2\{\sin^2\theta + \alpha^2\cos^2\theta + \alpha\rho\sin 2\theta\}.$$

Since $|\theta| < \pi/4$, $\cos^2\theta > \sin^2\theta$, so that $\sigma^2(Y_0) < \sigma^2(Z_0)$ if $\alpha > 1$ and $\rho\theta \geqslant 0$, $\sigma^2(Y_0) > \sigma^2(Z_0)$ if $\alpha < 1$ and $\rho\theta \leqslant 0$. Suppose first that $\alpha > 1$; then it follows from the formula for $\tan 2\theta$ that, whatever be ρ, $\rho\theta \geqslant 0$, so that $\sigma(Y_0) < \sigma(Z_0)$, and hence for any θ' with $|\theta'| \neq \pi/4$,

$$\sigma^2(Y_{\theta'}) = \sigma^2(Y_0)\cos^2\theta' + \sigma^2(Z_0)\sin^2\theta' < \sigma^2(Y_0)\sin^2\theta' + \sigma^2(Z_0)\cos^2\theta' = \sigma^2(Z_{\theta'}).$$

When $\alpha < 1$, $\rho\theta \leqslant 0$, so that $\sigma^2(Y_0) > \sigma^2(Z_0)$ and $\sigma^2(Y_{\theta'}) > \sigma^2(Z_{\theta'})$.

On the other hand, $\sigma^2(Z_{\theta'}) - \sigma^2(Y_{\theta'}) = \{\sigma^2(Z_0) - \sigma^2(Y_0)\}\cos 2\theta'$

vanishes if and only if $|\theta'| = \pi/4$.

In the intervals indicated under (c) we find

$$\rho' = \tfrac{1}{2}\{\sigma^2(Z_0)-\sigma^2(Y_0)\} \sin 2\theta'\Big[\big(\tfrac{1}{2}\{\sigma^2(Z_0)-\sigma^2(Y_0)\} \sin 2\theta'\big)^2+\sigma^2(Y_0)\sigma^2(Z_0)\Big]^{-\frac{1}{2}},$$

$$\sin 2\theta' = \frac{2\sigma(Y_0)\,\sigma(Z_0)}{\sigma^2(Z_0)-\sigma^2(Y_0)}\;\rho'(1-\rho'^2)^{-\frac{1}{2}},$$

where we can, if desired, express $\sigma(Y_0)$ and $\sigma(Z_0)$ in terms of $\sigma(Y_\theta)$, $\sigma(Z_\theta)$ and $\rho(Y_\theta, Z_\theta)$.

So $d\rho'/d\theta' = \{\sigma^2(Z_0)-\sigma^2(Y_0)\}\sigma^2(Y_0)\sigma^2(Z_0) \cos 2\theta'[(\frac{1}{2}\{\sigma^2(Z_0)-\sigma^2(Y_0)\} \sin 2\theta')^2+\sigma^2(Y_0)\sigma^2(Z_0)]^{-1\frac{1}{2}}$,

which, by (b), for $|\theta'| < \pi/4$ is positive when $\sigma(Y_\theta) < \sigma(Z_\theta)$ and negative when $\sigma(Y_\theta) > \sigma(Z_\theta)$, and which for $|\theta'| = \pi/4$ is zero; and

$d\theta'/d\rho' = (1-\rho'^2)^{-1} \operatorname{sgn} \{\sigma^2(Z_0)-\sigma^2(Y_0)\}\sigma(Y_0)\sigma(Z_0)\, [\{\sigma^2(Z_0)-\sigma^2(Y_0)\}^2-\{\sigma^2(Z_0)+\sigma^2(Y_0)\}^2\rho'^2]^{-\frac{1}{2}}$,

which for $|\rho'| < |\sigma^2(Z_0)-\sigma^2(Y_0)|/\{\sigma^2(Z_0)+\sigma^2(Y_0)\}$ is positive when $\sigma(Y_\theta) < \sigma(Z_\theta)$ and negative when $\sigma(Y_\theta) > \sigma(Z_\theta)$.

An immediate consequence of Theorem 3.1 and Theorem 3.2 (a) is :

Theorem 3.3: *Let the variances of Y_0 and Z_0 be finite and positive. Then, for $\theta \neq 0$, $\rho(Y_\theta, Z_\theta) = 0$ if and only if $\sigma(Y_0) = \sigma(Z_0)$.*

4. SYMMETRY

The following propositions connect relations between the distributions of Y_0 and Z_0 with relations between the distributions of Y_θ and Z_θ in case of symmetry. This is sometimes useful in applications; in any case, it seems natural that symmetry should play a role in the study of rotations.

Theorem 4.1: (a) *A necessary and sufficient condition for the existence of constants a and b such that the distributions of Y_0-a and Z_0-b be equal and symmetric about 0, is that Φ, the joint characteristic function of Y_θ and Z_θ, satisfy*

$$\Phi(t_1, t_2) = \Phi(-t_2, t_1) \exp \{(c \sin \theta+d \cos \theta)\, it_1+(c \cos \theta-d \sin \theta)it_2\},$$

where

$$c = a+b, \qquad d = a-b.$$

(b) *If there exist constants a and b such that the distributions of Y_0-a and Z_0-b are equal and symmetric about 0, the marginal distributions of*

$$Y_\theta-a \cos \theta-b \sin \theta \text{ and } Z_\theta+a \sin \theta-b \cos \theta$$

have the same property.

Proof: Let ϕ and ψ be the characteristic functions of Y_0 and Z_0, then

$$\phi(t) = \Phi(t \cos \theta, -t \sin \theta), \quad \psi(t) = \Phi(t \sin \theta, t \cos \theta).$$

If $\Phi(t_1, t_2) = \Phi(-t_2, t_1)$,

$$\phi(t) = \Phi(t \cos \theta, -t \sin \theta) = \Phi(t \sin), \theta, t \cos \theta) = \psi(t),$$

and $$\phi(t) = \Phi(t \cos \theta, -t \sin \theta) = \Phi(-t \sin \theta, -t \cos \theta) = \psi(-t) = \phi(-t).$$

Conversely, since $$\Phi(t_1, t_2) = \phi(t_1 \cos \theta - t_2 \sin \theta)\ \psi(t_1 \sin \theta + t_2 \cos \theta),$$

and consequently $$\Phi(-t_2, t_1) = \phi(-t_1 \sin \theta - t_2 \cos \theta)\ \psi(t_1 \cos \theta - t_2 \sin \theta),$$

we have, if $$\phi(t) = \psi(t) = \psi(-t),$$

$$\Phi(t_1, t_2) = \Phi(-t_2, t_1).$$

All other stated results are now immediate.

If Y_θ and Z_θ have the same normal distribution, but are correlated, $|\theta| = \pi/4$, but $(Y_0 - EY_0)/\sigma(Y_0)$ and $(Z_0 - EZ_0)/\sigma(Z_0)$ have the same normal distribution. For all other cases we are often (Konijn) interested only in the case of small $|\theta|$, so that the following proposition may be useful :

Theorem 4.2: *Let G and H be distributions for which the moment problem is determined, let* $F_\theta \in R[GH]$, *and* $|\theta| < \pi/4$. *If there exist constants a and b for which the marginal distributions of* $Y_\theta - a$ *and* $Z_\theta - b$ *are symmetrical about* 0 *and equal, then there exist* a_0 *and* b_0 *for which* $G(t-a_0) = H(t-b_0)$, *and G is symmetrical about* a_0.

Proof: $$EY_\theta - a = EZ_\theta - b = 0,$$

so if $$a_0 = a \cos \theta - b \sin \theta,$$

$$b_0 = a \sin \theta + b \cos \theta,$$

we have $$E(Y_0 - a_0) = E(Z_0 - b_0) = 0.$$

Suppose that for $$0 \leqslant h < k+1$$

$$E(Y_0 - a_0)^{2h} = E(Z_0 - b_0)^{2h},$$

$$E(Y_0 - a_0)^{2h+1} = E(Z_0 - b_0)^{2h+1} = 0.$$

Then the equality of

$$E(Y_\theta - a)^{2k+2} = \sum_{h=1}^{k+1} \binom{2k+2}{2h} E(Y_0 - a_0)^{2k+2-2h} E(Z_0 - b_0)^{2h} \cos^{2k+2-2h} \theta \sin^{2h} \theta$$

and $$E(Z_\theta - b)^{2k+2} = \sum_{h=1}^{k+1} \binom{2k+2}{2h} E(Z_0 - b_0)^{2k+2-2h} E(Y_0 - a_0)^{2h} \cos^{2k+2-2h} \theta \sin^{2h} \theta$$

yields $$\{E(Y_0-a_0)^{2k+2}-E(Z_0-b_0)^{2k+2}\}(\cos^{2k+2}\theta-\sin^{2k+2}\theta) = 0,$$

so, since $$|\theta| < \pi/4,$$

$$E(Y_0-a_0)^{2k+2} = E(Z_0-b_0)^{2k+2}.$$

Moreover, we obtain the expressions,

$$0 = E(Y_\theta-a)^{2k+3} = E(Y_0-a_0)^{2k+3}\cos^{2k+3}\theta+E(Z_0-b_0)^{2k+3}\sin^{2k+3}\theta,$$

$$0 = E(Z_\theta-b)^{2k+3} = -E(Y_0-a_0)^{2k+3}\sin^{2k+3}\theta+E(Z_0-b_0)^{2k+3}\cos^{2k+3}\theta,$$

which, since $\cos^{2(2k+3)}\theta+\sin^{2(2k+3)}\theta \neq 0$ (except in the case $\theta = 0$, when the result is immediate), gives

$$E(Y_0-a_0)^{2k+3} = E(Z_0-b_0)^{2k+3} = 0.$$

REFERENCE

KONIJN, H. S. : On the power of certain tests for independence in bivariate populations to appear in *Ann. Math. Stat.*

Paper received : May, 1955.

CAPITALIST EVOLUTION IN THE LIGHT OF KEYNESIAN ECONOMICS[1]

By NICHOLAS KALDOR

King's College, Cambridge

It was a great pleasure for me to receive an invitation to give a lecture to the Economics Faculty of the University of Peking. I feel greatly honoured by this invitation, which provides an opportunity for me to contribute to furthering the free communication of ideas among the different parts of the world. I am on a sabbatical year's leave from the University of Cambridge and since leaving England have already given a series of lectures at the University of Delhi and also lectures at some other Universities. After leaving China I shall deliver some lectures at the Waseda University of Tokyo, and before I return to Cambridge I shall also fulfil lecturing engagements at the Universities of Columbia and Harvard in the United States. The very variety of the ideologies prevailing in these different Universities provides a challenge to attempt to reduce differences in modes of thought by the common intellectual process of logical reasoning.

Nothing brings home to one so much the need for mutual understanding between different peoples as a journey around the globe by air. It shows how small our planet has become on which we all live. It is inevitable that for a long time to come the world shall be divided between two political systems, or two ways of life, of the Communist countries and of the capitalist countries, of the "old" and of the "new" democracies, or whichever other expression we may prefer to employ in this connection.

Nothing but good can come from attempts to learn to know and to understand each other's points of view. Whatever the differences, we simply cannot afford to be ignorant of each other's thoughts, ideas or situations. For it is a basic trait in the human character that ignorance breeds fear, and that fear breeds hostility. I am very grateful therefore to have been given this opportunity to promote the free communication of ideas. I am very fortunate also in having found such an admirable interpreter in Professor Hsu, a friend and a graduate of my own University. As an interpreter he may have to explain to you ideas with which he may not be in agreement. If so I hope he will express his own views in the subsequent discussion.

Considering how near we are to each other geographically it is remarkable how ignorant we are as regards each other's situation and prospects. In my own country there are many people who sincerely believe that the people of China labour under a sense of frustration and oppression, and live under extremely harsh economic conditions. Although I have only been here for a short time it did not take long to discover that all this is nonsense. Even if one cannot speak the language one can look into people's eyes; one can observe how people work, how they move around and how they behave generally; and it seems evident to me that the people of China do not feel oppressed and miserable — on the contrary, they are experiencing a great outburst of national energy and vitality, and they set about the task of economic development with zeal and enthusiasm. In fact, as far as I can see, nothing can stop China from becoming in a generation or so, one of the greatest powers, and from developing her human and material resources to the point where the standard of living of her 600 million inhavitants — who by that time will more likely number 1,000 millions — will approximate to that of the most advanced countries of the world.

---

[1] A lecture delivered at the University of Peking on 11th May, 1956.

The author is indebted to the Dean of the Department of Political Economy of Peking University for permission to reprint this lecture.

My purpose here today however is not to talk about China but about the capitalist countries. For my feeling is that the misconceptions prevalent among you concerning the state of affairs in capitalist countries are of the same order as those which exist in the capitalist countries concerning China. I may be wrong, but I certainly received the impression since I arrived here that it is widely believed that in capitalist countries there is stagnation (or worse) and not progress; that the standard of living of the masses is falling and not rising, and that the capitalist countries are inevitably heading for some tremendous economic crisis —that the capitalist system of the Western countries must inevitably collapse sooner or later under the weight of its own inherent contradictions.

If these views are held, I am sure that they are false. In the successful capitalist countries of Western Europe and North America — in the countries that is to say, where the capitalist system has successfully replaced the precapitalistic modes of production—the total level of production, the general standard of living of the workers is rising year by year, and has been rising for at least the last hundred years or so. In fact the rate of improvement of the last decade was at least as great, if not greater, than the average rate of improvement in the previous nine decades. It is true that in a capitalist economy economic progress though continuous, does not occur in a smooth manner from year to year—it is subject to fits and starts, to periodic fluctuations known as the trade cycle. But the most important feature of capitalism which distinguishes it from pre-capitalistic societies is its technical dynamism, the continuous improvement in methods of production as against the unchanging techniques of peasant cultivation and handicraft production. Not all countries experience this process of a successful capitalist evolution—in many areas of the world, like South-East Europe, the Middle East, most of Asia and the greater part of Africa and Latin America, capitalist development has never got properly going and the capitalist sector remained confined to a small segment of the economy. It is certainly a most intriguing subject for inquiry (but one which cannot be gone into here) just why the capitalist mode of production spread so fast in some societies and not in others.

It is true that even in countries where the stage of capitalism has successfully established itself, capitalism does not necessarily secure the full employment of labour, even though it does tend to raise the earnings of workers in proportion to the rise in output per man. In fact the proportionate rise in productivity and in real wages, the historical constancy of the relative shares of capital and labour, is one of the remarkable features of capitalist evolution. Contrary to Marx's own analysis, in developed capitalist societies the real earnings of workers tend to rise automatically with output per head. This rise in the standard of living of the working classes does not in itself imply a reduction in economic inequality: for the richesse of the biggest capitalists rises at least as much as the general standard of living. There is no automatic tendency towards a more equal distribution in the ownership of property in the course of economic progress.

Unemployment, fluctuations, and growing concentration in the ownership of property, are not, however, in my opinion, inevitable features of capitalist evolution. Thanks to the work of some economists and notably to Keynes, we know far more about the mechanics of capitalist evolution than we did even a decade or two ago; and we are now in a position to mould it, by suitable public controls, according to a desired pattern. Western Socialists like myself believe that men can control the endogenous forces of human society in much the same way as through science we can control the forces of nature. We believe that with suitable controls we can secure continuing full employment, a steady development of productive forces and the gradual reduction in economic inequality at the same time—without any sudden or revolutionary change in social and political institutions which could be viewed as a liquidation of capitalism.

The view that the progress of capitalism involves ever widening economic crises and that the system must finally collapse was not of course derived from empirical observation—it could not have been. It is based on the theoretical analysis of the workings of capitalism by Marx, which was elaborated mainly in the sixties and seventies of the last century. Perhaps I ought to make it clear at the start that I myself regard Marx's economic theory—which itself was a further development of Ricardo's model — as an extremely powerful tool for analyzing the mechanics of capitalism at the stage of transition from a pre-capitalist to a capitalist society. Where Marx and I differ—if I may use such an expression without undue irreverence — concerns the latter evolution of capitalism, after the transition from the pre-capitalist to the capitalist stage had been completed. In my view Marx's predictions concerning this were mistaken in some very important respects. It is true that one of his predictions—concerning the increasing concentration of production in the hands of a few large concerns—proved accurate. But his associated prediction that this process involved the steady crystallization of society between a small class of exploiters and the vast mass of the exploited, accompanied by the steady worsening of the living conditions of the working classes – the so-called "immiseration of the proletariat"—proved wholly inaccurate. The standard of living of the proletariat in the course of capitalist evolution has not fallen but has substantially risen. In countries like the United States or Britain, or the Scandinavian countries it has trebled or even quadrupled since Marx's day — especially if hours of work as well as earnings are taken into account. It is no accident that the revolution leading to the dictatorship of the proletariat has come about not, as Marx predicted, in the countries of the most advanced stage of capitalism, but in countries where the capitalist mode of production has not properly established itself. On Marx's theory it should have been inevitable that the proletarian revolution should first arise in the countries of Western Europe and North America, and not in Russia or China.

Just what is the explanation for these developments ? It is not sufficient to appeal to historical facts in order to refute a theoretical scheme. We want to know why things happen in a certain way and why they do not happen in some other way—in the way in which Marx predicted them. Without a theoretical scheme which is capable of explaining historical developments we are merely groping in the dark; and are unable to assess the true significance of the factual criticisms that can be brought against the Marxian analysis.

My purpose in this lecture therefore is not to criticize Marx but to present to you an alternative theoretical scheme concerning the laws of evolution of capitalism which is based on Keynesian economics. In order to show just where the important difference between the two kinds of theory lies, it is best, I think, to start with the Marxian scheme with which you are already familiar and to focus attention on the precise points at which the two theories part company.

Let me therefore start with a resumé of Marxian theory. I think the three most important features of Marx's theory of capitalism are as follows :

(i) that the wages of labour are determined by the cost of reproduction of labour—the necessary minimum subsistence of the labourer—while the surplus of production over this minimum or subsistence-consumption accrues to the capitalist in the form of profit. Profit is thus a residue between output per man and minimum consumption per man;

(ii) that the supply of wage labour in the market always exceeds the demand for wage labour; the excess being the reserve army of labour which is essential to the functioning of capitalism;

(iii) that as a matter of competitive necessity (at any rate in the competitive stage of capitalism) the profits of the capitalist are largely re-invested or devoted to accumulation for a capitalist who does not continually re-invest his profits and thereby expand the scale of his business, will fall by the wayside in the competitive struggle.

The share of profits in output is thus determined by "surplus value"— the difference between the product of labour and the cost of labour. Thus if a worker required four hours work to provide for his own subsistence and the length of the working day is 10 hours, the ratio of profits to wages is as 6 : 4. Or, writing P for profits, Y for total income (or production), C for the subsistence cost of the labourer, L for the labour force, and SV for surplus value,

$$SV = Y - CL$$

$$\frac{P}{Y} = \frac{SV}{SV+CL} \qquad \ldots \quad (1)$$

CL being identical with the "variable capital" of the community.

Marx regarded the co-existence of (ii) and (iii) as one of the basic contradictions of the capitalist system—the search for more profit destroying the basis on which the profit system is built. This is because the demand for wage labour depends on the accumulation of capital. As capitalist enterprise expands through accumulation, both the demand for labour and the supply of wage labour expands, the latter on account of the break-up of pre-capitalistic units of production. But in time as accumulation proceeds faster and faster, the demand for wage-labour must also increase faster and faster; hence it must sooner or later over-take the increase in supply, and thereby extinguish the "reserve army". When this happens, wages rise and profits fall since the factor which previously tied wages to the subsistence level—the excess of job-seekers over the number of jobs available—is no longer present. When the capitalists are obliged to bid against one another in order to obtain labour, wages rise and profits are wiped out. This causes a crisis which goes on until the reserve army is restored again through the adoption of more labour-saving methods of production —through a "higher organic composition" of capital. The existence of a reserve army is thus essential to the preservation of profits. If the demand for labour exceeds the supply of wage labour there is nothing in Marx's theory to prevent wages from rising until profits are wiped out altogether[1].

It is at this point that Keynes' analysis leads to fundamentally different conclusions. Granted that in times of labour scarcity—and particularly when different employers are in keen competition with one another—wages must rise in money terms, this does not mean that they will rise in *real* terms—in other words, that the rise in money wages means an equivalent reduction in profits. It is, in my view, one of the chief merits of Keynes' analysis to have shown that money wages and real wages (or wages expressed in terms of the product, or as a share of the product) are determined by fundamentally different conditions. It is only the level of money wages which is directly influenced by the relative scarcity or abundance of labour in the labour market. The level of *real* wages is determined by quite different forces and in times of labour scarcity (or full employment) it must be determined by the condition that the total demand for commodities of all kinds must neither exceed nor fall short of the total supply of commodities of all kinds. Or, to put the same proposition in different words: real wages must be such as to make total expenditure of both the capitalists and the workers neither greater nor less than the total supply of goods available to meet that expenditure. In a sense capitalists and workers compete for the available commodities; but while the workers' expenditure is mostly on consumption, the capitalists' expenditure is partly on consumption and partly on business investment, or accumulation. Capitalists are in an advantageous position in that they posses purchasing power which is large relative to their expenditure in a period, whereas the workers' reserve of purchasing power is small or non-existent. The capitalists' expenditure may therfore be entirely independent of their current receipts or earning—at any rate the connection between the two is far less rigid or direct than is the case with the income and expenditure of the workers. Workers always spend more if they earn more, and vice-versa, are forced to spend less when they earn less. Hence when wages rise in real terms (and profits correspondingly fall) the total demand for goods rises— since the workers increase their *real* expenditure side by side with the increase in their income, whilst the capitalists' expenditure is not automatically reduced with the fall in their real earnings. Similarly when wages fall in real terms and profits rise, the total demand for goods in *real* terms is reduced since the reduction in the workers' demand for commodities is not automatically compensated by an equivalent increase of the capitalist's demand for goods — since the latter was in any case not confined or limited by their current earnings. It can be easily seen therefore, that in any given situation there must be some division of the product between wages and profits which makes the total demand for commodities equal to the total supply. If wages are higher than is indicated by this division, demand will exceed the supply, and prices will inevitably rise in relation to wages, and thereby reduce wages as a share of output or income. Similarly if wages were lower than this, the demand would be short of supply, prices would tend to fall and the share of wages in output would rise. Under conditions of full employment the relation of prices and wages must always be such as to prevent either an excess or a deficiency of total demand over total supply. Another way of putting the same proposition is to say that since capitalists' ependiture is (relatively) independent of their current earnings whilst the worker's expenditure depends on them, capitalists as a class will in fact earn what they spend, whilst the workers spend what they earn.

---

[1] Adam Smith in the *Wealth of Nations* came near to stating an essentially similar conclusion. Cf. Book I, Ch. VIII.

The capitalist's expenditure as mentioned before is of two kinds: expenditure on investment (for purposes of business expansion) and expenditure on personal consumption. Keynes' analysis of the multiplier has shown total income as the outcome of two factors, investment outlay and the propensity to save. The same analysis can be applied to the determination of total or real income and output if the distribution of income between wages and profits is treated as given, or to the distribution of income between wages andp rofits, if total real income or output is treated as given. If we assume for simplicity that the working classes' expenditure is equal to their income (i.e., that they neither accumulate property through current savings nor draw on past savings to supplement their expenditure on consumption out of current earnings) the determination of the share of wages reduces to the simple formula:

$$\frac{P}{Y} = \frac{1}{1-c}\,\frac{I}{Y} \qquad \dots \quad (2)$$

where the terms : P Y and I stand for profits, income and investment expenditure respectively and c for the proportion of income spent on consumption. Since the sum of consumption and saving equals income, this formula can also be written in the simpler form:

$$\frac{P}{Y} = \frac{1}{s}\,\frac{I}{Y} \qquad \dots \quad (2a)$$

where s stands for the proportion if income saved.[1]

On this analysis the share of profits (and thus the share of wages) in income is determined once the ratio of investment to output, and the capitalists' propensity to consume (i.e., capitalists' consumption as a proportion of their income), are known. For the ratio of investment to output we have however yet another equation :

$$\frac{I}{Y} = Gv \qquad \dots \quad (3)$$

where G stands for the (average) expected rate of market expansion (of businessmen) and v stands for the ratio of capital to output, i.e., the amount of investment required per unit of output capacity, expressed as a multiple of annual output.

---

[1] If the workers' savings are not zero, but are either positive or negative the formula is more complicated without however making any essential difference to the analysis so long as the savings of workers, expressed as a percentage of wages are small in relation to capitalists' savings as a percentage of profits. I use the terms "profits" and "wages" in an inclusive sense, to embrace all forms of income from property and from work, respectively. There are some forms of income from property which are not the profits of enterprise (such as interest and rent) and which are best treated as a deduction from the total profits of enterprises. Similarly not all incomes from work are wages: there are, e.g. the salaries of the executive personnel and professional earnings. So long however as the category of mixed incomes (where the element of income from work and from property is of equal weight) is relatively unimportant, no significant error is introduced by the simplification of regarding all incomes as belonging to one of these two categories. (From the point of view of the particular theory that we are discussing the most important difference between property incomes and wage incomes lies in the fact that in the one case the total purchasing power at a person's command is greatly in excess of his total expenditure in a particular period, whilst in the other case it is not.)

In other words assuming that the technical relationship between the value of capital and output— the amount of capital required per unit of output— and the average rate of expansion of markets which businesses expect are known, the ratio of investment to output is also determined. This theory therefore asserts that the share of profits depends fundamentally on three things: on how much investment is required in order to expand output capacity by any given amount (v) ; the expectations of entrepreneurs concerning the growth of sales (or markets) which govern the planned rate of expansion of output capacity (G); and finally on the propensity of consumption of capitalists (c) which governs their consumption in relation to their income, and hence ultimately to their investment expenditure. Thus if $v = 4$ (i.e., the capital investment required to produce output is four times the annual output), and if $G = 3$ per cent per annum (i.e., the average expectation of entrepreneurs is that the markets expand at the rate of 3 per cent per annum), and if $c = 50$ per cent (if capitalists, on the average, consume one-half of their profits) I will be 12 per cent ($4 \times 3\%$), and P according to equation (2) above, 24 per cent. If we assume that $c = 66\ 2/3$ per cent, (i.e., capitalists consume two-thirds of their profits), P (the share of profits) would become 36 per cent ($3 \times 12\%$) and the share of wages 64 per cent.

The main factor which requires further explanation in this theoretical scheme is the expected rate of expansion of markets, G. We know that quite apart from expectations, there is in any given situation a certain maximum potential rate of expansion of production, determined by the growth of working population and the rate of increase in productivity per worker. Let us call this expansion rate, G' as determined by the formula:

$$G' = t+p \qquad \dots (4)$$

where t stands for technical progress, as measured by the annual increase in productivity per man and p for the increase in population, also expressed as an annual percentage[1]. The condition of a smoothly developing economy is that :

$$G = G'$$

hence

$$\frac{I}{Y} = Gv = (t+p)v \qquad \dots (5)$$

There is no necessity of course for this equality to prevail over short periods, though the two tend to approximate over long periods through long term adjustments of G' to G, as well as adjustments of G to G'.

As the relations here are rather complicated let me assume to start with that t and p are constant factors determined independently of the endogenous forces of the economy. In that case the condition of a successfully developing capitalist economy is that

$$G \geqslant G'$$

In other words that entrepreneurs should expect markets to expand by at least as high a rate or some higher rate than the maximum technically possible rate of expansion of the economy If G exceeds G' (which we may regard as the normal feature of successful capitalist economies) investment will tend to be excessive in the sense that the capacity to produce output will tend to expand faster than output itself—leading sooner or later to the appearance of excess capacity, which in turn leads to a temporary interruption of the investment process. This I regard as the main reason why economic progress should proceed in fits and starts (through cyclical fluctuations of "booms" and "slumps") in capitalist societies.[1]

---

[1] The formula $G' = t+p$ is only an approximation. Th exact formula is $G' = (1+t)(1+p)$, which is approximately equal to $t+p$ when t and p are small fractions.

If on the other hand $G < G'$, the economy will not succeed in growing at its natural rate over longer periods, but the actual growth will be insufficient to prevent growing unemployment. If $G = 0$, the economy will relapse into a state of complete stagnation.

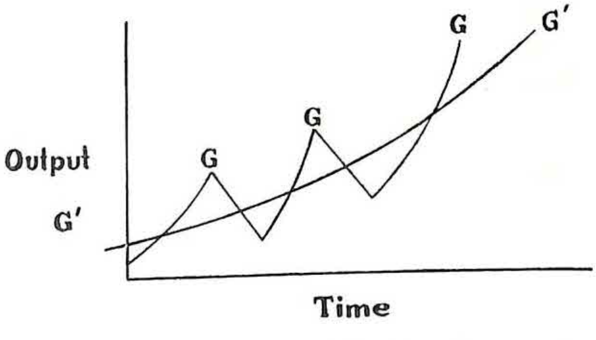
A capitalist economy is therefore exposed to two different hazards. If G exceeds G' (especially if it exceeds it by a considerable margin) the economy will be subject to violent booms and slumps, with all the economic and social instability which this entails. If on the other hand G falls short of G' (especially if it does so by a considerable margin) the economy will relapse into stagnation which must ultimately bring down t and p, and thus G'.

And here I would like to introduce a further complication. Whilst over shorter periods it is G which tends to adjust itself to G' (through the relative length of booms and slumps) over longer periods it tends to be the other way round: technical progress and population growth both tend to be accelerated, or slowed down (as the case may be) by the endogenous forces of expansion in the economic system. When G is large, capital accumulates at a faster rate than the working population, a situation of labour-scarcity must sooner or later be reached, which stimulates the introduction of labour-saving techniques; at the same time the rapid growth in the supply of consumer goods also stimulates the growth in population. The opposite happens when G is small—which may lead to a state where both G and G' tend towards zero.

There is therefore no inevitability about economic progress in a capitalist economy (or for that matter, in a socialist economy) it all depends on whether those in charge of production have the incentives and the will to pursue a vigorous expansion of output capacity. It is this expansion of output capacity which through its indirect consequences on the demand for labour and on the supply of consumable goods stimulates both t and p, and thus brings about the basic conditions for a sustained expansion over longer periods.

---

[1] If G' is given, the expansion rate G' can be indicated by a logarithmic curve, in a diagram showing time horizontally and output on the vertical axis. If $G > G'$, then the slope of the segment of the logarithmic curve indicating G must exceed, at any given time, the slope of the curve indicating G'. Since however, the economy cannot, over longer periods exceed the rate of expansion G', the actual progress of the economy must proceed by jerks, as shown in the diagram below.

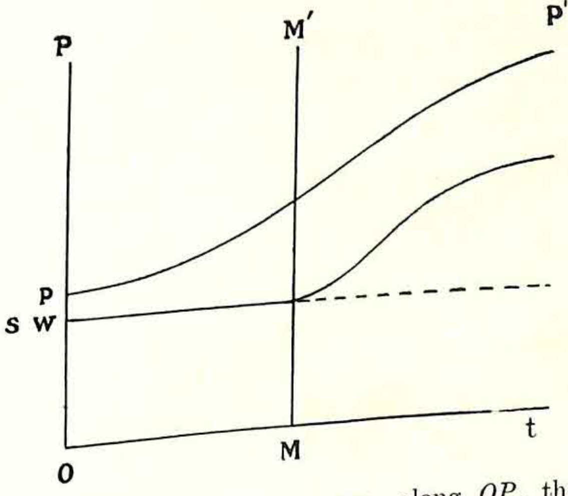


G' is also part of a single logarithmic growth curve but it is discontinuous due to the necessity to keep the actual expansion in harmony with G.

It remains to show how these two theoretical models, the Marxian and the Keynesian, are related to each other. It can be shown that either may be operative, depending on circumstances. Thus if the share of profit in output as indicated by the Keynes formula (equation (2) above) is higher (and the share of wages lower) than that indicated by the Marx formula (equation (1) above), the Marxian formula will be operative and the Keynesian inoperative, and vice-versa. Thus the Marxian formula indicates the minimum limit below which the share of wages cannot fall, irrespective of output per head, and the Keynes formula indicates the maximum above which the share of wages cannot rise irrespective of the scarcity or superabundance of labour. In general that particular formula will operate which yields higher real wages per head.

Thus supposing, as before that the length of the working day is 10 hours, and the share of profits indicated by the Keynes formula is 50 per cent (as would be the case if we assumed, e.g., that $v = 5$, $G = 2.5\%$ per annum and $c = 80\%$) so that the wages indicated by the formula are the equivalent of 5 hours of work. If the amount of labour necessary for the worker's subsistence is 6 hours, and in consequence, the surplus value is only 4 hours, profits will be compressed below the amount indicated by the Keynes formula, i.e., to 40 per cent. But if on the other hand, the product-equivalent of the workers' cost of subsistence takes only 4 hours of work, and the surplus value is 6 hours, profits will still be only 50 per cent, as indicated by the Keynes formula, and the real wage of the workers will exceed the cost of reproduction of labour by the equivalent of 1 hours' work —i.e., the capitalists will only get 5/6 of the surplus value, and the workers will get 1/6.

In the early stages of capitalist development, when productivity per man is relatively low, the surplus value is likely to be considerably less than that necessary to satisfy equation (2). During that period the Marxian scheme operates: wages remain at subsistence levels, despite rising productivity per man. But as productivity and surplus value rise, a point must be reached sooner or later, when the surplus value equals or exceeds the profit indicated by the Keynes formula: from then onwards the share of profits ceases to rise and real wages begin to rise, pari passu with the rise in productivity, as shown in the following diagram :



Measuring time along Ot, the product and the wage along OP, the curve $p\text{-}p'$ indicates the growth in productivity per man over time; $S\ W$ indicates sub-sistence wages. At some critical point $(M—M')$ the product per man becomes high enough for the surplus value

]

($=p-s\,w$) to equal or exceed the profits shown by equation (2). Henceforth real wages are no longer tied to the subsistence level; and, apart from changes reflecting changes in the variables in equations (2) and (3), the share of wages in output remains constant.

Thus the Marxian scheme is applicable in the early stages, and the Keynesian scheme in the later stages of capitalist development. This explains why in the early stages of development the wage earners obtain so little benefit from the growth in production whilst in the later stages real wages rise so much above the subsistence level and go on rising with further increases in productivity.

What I have said in this lecture has only been intended as the barest outline of a theory of capitalist economic development. There are numerous gaps to be filled in, and complications to be introduced which it is impossible to do in a single lecture. I should, however, draw attention to one important complication; equations (2) and (3) are to be interpreted as "long-run" equations, as the determinants of the "normal" share of profits and of wages. Short-period fluctuations in the variables G, v, c may not reflect themselves in corresponding variations in the share of wages and of profits, but either in fluctuations in output or in investment-rationing. This is because in the short period there are elements of resistance against downward revisions either in the share of profits or in the share of wages, and this tends to stabilize the shares around the customary level. Thus if there is a sudden fall in Gv, and hence $\frac{I}{Y}$ equation (2) tells us that the margin of profits on turnover is correspondingly reduced, and the fall in investment demand by entrepreneurs is compensated for through an equivalent induced increase in consumption demand by wage earners (due to the fall in prices relative to wages). In fact, however, profit margins may not be cut, or only tardily, with the result that aggregate real demand, and hence aggregate production, is reduced. In the opposite case where there is a sudden rise in Gv, and hence in $\frac{I}{Y}$, equation (2) indicates a rise in prices relative to wages which deprives the workers of purchasing power and thus reduces their consumption so as to release resources for the rise in investment demand. In fact however the workers resist this cut in their accustomed standard of living through demanding higher money wages, which in turn leads to further increases in prices and of money wages, (and thus to a process of inflation) until the Government takes steps (in order to protect the currency) to "ration" investment through monetary controls, or licenses, etc., and thereby cut $\frac{I}{Y}$ below Gv. These elements of resistance or inertia serve to reinforce the long-term stability in the shares of pofits and wages due to the factors mentioned earlier.

In conclusion I should like to emphasize again what I said at the beginning. There is no inherent necessity in a capitalist economy or any other economy of sustained evolution. There may be stagnation rather than development; progress may take the form of violent ups and downs, and not of a smooth or steady process; progress may be attended by a growing concentration of wealth and economic power in the hands of a few individuals. But none of these tendencies if they arise are either inevitable or unavoidable. They are all subject to social control once we understand the manner of operation of economic and social forces. It is my belief that in a progressive social democracy they could all be avoided.

Paper received : August, 1956.

THE FOUNDATIONS OF STATISTICS*

By P. C. MAHALANOBIS

Indian Statistical Institute, Calcutta

INTRODUCTION

0.1. I am particularly grateful to Professor F. Gonseth and Professor A Linder, the organizers of the Zürich Symposium which was held in the Swiss Federal Institute of Technology in Zürich in April 1953, for having invited me to open the discussion on the foundations of statistics. In my opening observations I had referred to certain ideas in the Indian-Jaina theory of *syādvāda*[1] which aroused much interest; and I was requested to give a fuller account of Jaina views at the time of preparing the text of my communication. I am doing this in Part I of the present paper.

0.2. There was also a brisk discussion on some of the views which I had placed before the symposium. It was suggested that I should take into consideration these comments and amplify to some extent my own arguments in writing out my communication. I intend to give my own views in Part 2 of the present paper broadly on the basis of the observations made by me at the Zürich symposium with, however, certain additions and amplifications to deal with some of the comments made by the other speakers.

0.3. I wish to offer my sincere thanks to Professor Linder for looking through the manuscript of the paper and for making valuable comments and suggestions.

PART 1. THE INDIAN-JAINA DIALECTIC OF SYĀDVĀDA IN RELATION TO PROBABILITY

1. BRIEF HISTORY OF SYĀDVĀDA

1.1. There are certain ideas in Indian-Jaina logic called *syādvāda* which seem to have close relevance to the concepts of probability, and which can, therefore supply a convenient background to my own observations on the foundations of statistics. It is always difficult to be sure about the exact meaning of logical and philosophical phrases which were current 1500 or 2500 years ago: and it is not claimed (and I also agree that it would not be

---

*[This paper originally appeared in *Dialectica* Volume 8, No. 2, 15 June 1954 (Switzerland) and is being reprinted with the permission of the *Dialectica* —Editor.]

[1] Usually described as "The dialectic of the sevenfold categories of knowledge". In the present paper I have given references to books in the English language and not to original Sanskrit texts because these are not likely to be readily available in Europe. The English books (to which I have referred) give very full bibliographical references to original texts.

correct to claim) that the concept of probability in its present form was recognized in *syādvāda* but the phrases used in *syādvāda* seem to have a special significance in connexion with the logic of statistical inference.

1.2. I shall first give a brief historical account of *syādvāda*. Jaina religion and philosophy came into prominence from the time of its great leader Mahāvira (599-527 B.C.) who was a contemporary of Buddha, the founder of the Buddhist religion. The earliest reference to *syādvāda* occurs in the writings of Bhadrabāhu who is believed to have given the following explanation of *syādvāda*: syāt = "may be", and *vāda* = "assertion", or the assertion of possibilities.[1]

"The *syādvāda* is set forth as follows: (1) May be, it is; (2) may be, it is not; (3) may be, it is and it is not; (4) may be, it is indescribable; (5) may be, it is and yet is indescribable; (6) may be, it is not and it is also indescribable; (7) may be, it is and it is not and it is also indescribable."[2]

1.3. There were two authors of the name Bhadrabāhu, the senior belonging to the period 433-357 B.C., and the junior to about 375 A.D., and it is not definitely known whether the above explanation was given by the senior or the junior Bhadrabāhu; but the above exposition is usually ascribed to the senior Bhadrabāhu of the 4th century B.C..[3] There is indisputable mention of *syādvāda* in the *Nyāyāvatāra* of Siddhasena Divākara[4] (about 480-550 A.D.). A little later Samantabhadra (about 600 A.D.) gives a full exposition of the seven parts of *Syād-vāda* or *Sapta-bhaṅginaya* in his *Āptamimāṁsā*.[5] It is clear that *syādvāda* was well developed by the sixth century A.D., and received a great deal of attention in the mediaeval period of Indian logic; the *syādvādamañjarī* of Mallisena (1292 A.D.), for example, is a separate treatise on the same theory.[6] There are, of course, still later works such as Vimala Dāsa's *Saptabhaṅgītaraṅgiṇī* and a large number of mediaeval and modern commentaries. I am, therefore, dealing with a well-known theme which is considered to be the most original contribution of Jaina logic to Indian thought.[7]

---

[1] Satis Chandra Vidyabhusana, *A History of Indian Logic* (Calcutta University 1921), pp. 167. (This book will be referred to as *HIL*.)

[2] *HIL*, pp. 167-168.

[3] *HIL*, p. 167.

[4] *HIL*, p. 181: "It is the perfect knowledge of things taken from all possible standpoints. Thus a thing may be, may not be, both may or may not be, etc. according as we take it from one or the other standpoint."

[5] *HIL*. pp. 182-184.

[6] Jadunath Sinha, *History of Indian Philosophy* (Central Book Agency, Calcutta, 1952). Vol. II, p. 179. (This book will be referred to as J. Sinha, *HIP*).

[7] Satkari MOOKERJEE: *The Jaina Philosophy of Non-Absolutism* (Bhāratī Jaina Pariṣat, Calcutta, 1944), p. 191. (This book will be referred to as *JPN*).

2. DIALECTIC OF SEVEN-FOLD PREDICATION

2.1. I shall next refer to the actual text in Sanskrit of the dialectic of sevenfold predication (*saptabhaṅgīnaya*):

(1) *syādasti*[1] = may be, it is.
(2) *syātnāsti* = may be, it is not.
(3) *syādasti*[1] *nāsti*[2] *ca* = may be, it is, it is not.
(4) *syādavaktavyaḥ*[3] = may be, it is indeterminate.
(5) *syādasti ca*[4] *avaktavyaśca*[5] = may be, it is and also indeterminate.
(6) *syātnāsti ca avaktavyaśca* = may be, it is not and also indeterminate.
(7) *syādasti nāsti ca avaktavyaśca* = may be, it is and it is not and also indeterminate.

2.2. The word *syāt* has been translated as "may be" but this does not bring out the full implications. The Sanskrit word in mentioning one possibility has also some indirect allusion to other possibilities. The Sanskrit word *asti* may be rendered as "it is", "it exists", or "it is existent"; and *nāsti* is the negation, i.e. "it is not", "it does not exist", or "it is non-existent". The third category predicates the possibility of both *asti* and *nāsti*; of both "it is" and "it is not". The first three categories conform thus to the categories of classical logic and do not present any difficulty.

2.3. The fourth category is *avaktavya* which I have translated as "indeterminate". Other authors have used the words "indescribable",[6] or "inexpressible" or "indefinite". For example, Satkari Mookerjee explains "The inexpressible may be called indefinite"... (*JPN*, p. 115) I prefer "indeterminate" because this is nearer the interpretation which I have in mind.

2.4. It will be useful if at this stage I give an illustration. Consider the tossing of a coin; and suppose it turns up "head". We may then say (1) "it is head" (now). This also implies, (2) "it is not-head" (on some other occasion). The third category follows without difficulty, (3) "it is, and it is not" which is a synthetic predication based on both (1) and (2). The fourth category predicates that the position is still (4) indeterminate.

2.5. This, however, does not exhaust the possibilities of predication or modes of knowledge. For example, if we know that it is a coin which has "head" on one side and

---

[1] The two words *syāt* (may be) and *asti* (it is) are compounded (by rules of pronunciation) in one compound phrase *syādasti*.

[2] *nāsti* = not-is or it is not.

[3] The compound-phrase consists of two words *syāt* (may be) and *avaktavyaḥ* (or inexpressible, or indeterminate).

[4] *ca* = and or also.

[5] By rules of pronunciation the two words *avaktavyaḥ* and *ca* are compounded into *avaktavyaśca*.

[6] For example, Satis Chandra VIDYABHUSANA in *HIL* and other works.

"not-head" or "tail" on the other side, and we also know that it must turn up either "head" or "tail", we may then predicate that (5) there exists one type of indeterminateness which is capable of being resolved in terms of the first four categories. On the other hand we may know that the subject of discourse is not a coin but something else to which the category of indetermination in the above sense cannot apply, we may then use the sixth mode of predication and assert that (6) there does not exist that type of indeterminateness which is capable of being resolved in terms of the first four categories. Finally, there is the seventh mode of knowledge where we may be able to predicate that sometimes the possibility of resolution of indetermination exists (as in the fifth mode) and sometimes this possibility does not exist (as in the sixth mode).

2.6. According to *syādvāda*, the above seven categories are necessary and are also sufficient so that they exhaust the possibilities of knowledge. There is a minority view which holds that there are further possibilities of (8) *vaktavyaśca avaktavyaśca*, a kind of duplicated indeterminateness together with successive categories of the fifth, sixth, and seventh types in an infinite regression but the accepted opinion is that the hypothetical eighth category is identical with the fourth so that there is no need of more than seven categories.

2.7. I should like to emphasize that the fourth category is a synthesis of three basic modes of "it is" (assertion), "it is not" (negation), and inexpressible, or indefinite, or "indeterminate" (which itself is resolvable into either "it is" or "it is not"), and supplies the logical foundations of the modern concept of probability. Consider the throw of a coin. It has the possibility of head (it is) or not-head (it is not); sometimes head and sometimes not-head; and the combination of both possibilities of "it is" and "it is not" in an yet indefinite or indeterminate form. The fifth category of knowledge in Jaina logic predicates the existence of indetermination (which we may perhaps interpret, in modern language, as the assertion of the existence of a probability field). The sixth category denies the existence of a probability field; while the seventh category covers the whole range of possibilities mentioned in the other six categories.

3. RELATIVISM

3.1. It would be of interest to consider some further aspects of Jaina logic. The points to be stressed are that Jaina thought is non-abosolutist (that is, it is relativist) and realist. Siddhasena Divākara (480-550 A.D.) in *Nyāyāvatāra* (which is accepted as the earliest Jaina work on pure logic at present available) gave an exposition of *syādvāda* (knowledge of the all-sided method) of which the authentic text is as described below:

"*Syādvāda*, which literally signifies ascertion of possibilities, seeks to ascertain the meaning of things from all possible standpoints. Things are neither existent nor non-existent absolutely...*Syād* which signifies "may be" denotes all these seven possibilities, that is, a thing may be looked at from one of the above seven points of view, there being no eighth alternative."[1]

---

[1] *Nyāyāvatāra* edited by Satis Chandra VIDYABHUSANA. (Indian Research Society, Calcutta, 1909), pp. 29-30.

3.2. It has been pointed out that:

"All objects are multiform (*anekānta*) according to him (i.e. the Jania). From their many-sided nature it follows that all judgments are relative. They are true under certain conditions. They are conditional or hypothetical. No judgments are absolutely true. The word "perhaps" must be added to all judgments to indicate their conditional character. This is *Syādvāda* or the doctrine of relativity of judgments."[1]

"The Jainas emphasize manifold nature of real things which are endowed with infinite qualities, modes, and relations to the other things.[2] They have identity-in-difference. The Vedantists emphasize pure identity and deny plurality. The Jainas emphasize manifoldness of inter-related reals and deny pure identity. They are anti-Absolutists. They are advocates of relative pluralism."[3]

3.3. It has been also pointed out that:

"Thus the Jainas hold that no affirmation, or judgment, is absolute in nature, each is true in its own limited sense only, and for each one of them any of the above seven alternatives (technically called *saptabhaṅgī*) holds good. (See *syādvādamañjarī* with Hemachandra's commentary p. 166 etc). The Jainas say that other Indian systems each from its own point of view asserts itself to be the absolute and the only point of view. They do not perceive that the nature of reality is such that the truth of any assertion is merely conditional, and holds good only in certain conditions, circumstances, or senses (*upādhi*). It is thus impossible to make any affirmation which is universally and absolutely valid. For a contrary or contradictory affirmation will always be found to hold good of any judgment in some sense or other. As all reality is partly permanent and partly exposed to change in the form of losing and gaining old and new qualities, and is thus relatively permanent and changeful, so all our affirmations regarding truth are also only relatively valid and invalid. Being, non-being and indefinite, the three categories of logic, are all equally available in some sense or other in all their permutations for any and every kind of judgment. There is no universal and absolute position or negation, and all judgments are valid only conditionally."[4]

4. REALISM

4.1. Jaina logic is essentially realistic: "The Jaina philosopher maintains that existents are possessed of an infinite number of attributes and characteristics which can be discovered by experience alone. ...He refuses to put a premium on internal intuition. The mind, even with its active contributions, which the Jaina does not seek to deny, is believed by him to be an instrument of discovery and not a creator of facts." (*JPN*, p. 1)

---

[1] J. SINHA, *HIP*, vol. II, 1952, pp. 205-206.

[2] It is worth noting that the Jaina view in this respect has much similarity to A.N. WHITEHEAD's "inexhaustibility of nature". Also cf. V.I. LENIN: "Materialism and Empiriocriticism."

[3] J. SINHA, *HIP*, vol. II, 1952, pp. 208.

[4] S. DASGUPTA, *History of Indian Philosophy*, vol. I, pp. 180-181 (Cambridge University Press, 1922).

"Logic has to work upon the data of experience and is as much an instrument as experience is." (*JPN*, p. 5)...."Pure logic, prior to and independent of experience, is a blind guide to the determination of truth. Logic is to rationalize and systematize what experience offers." (*JPN*, p. 78).

"A thing is existent, is non-existent and is both existent and non-existent, but always subject to limitations imposed by objective differences of substance, time, space and attributes (*dravya-kṣetra-kāla-bhāvapekṣayā*).[1] The differences in predication are not due to our subjective contemplation from different angles of vision, but are founded upon objetively real attributes. They are facts irrespective of the consideration whether we contemplate them or not." (*JPN*, p. 107.)

4.2. "The Jaina does not see any reason why things should be particulars alone. Things are, according to the Jaina, both universals and particulars together.... A real is a particular which possesses a generic attribute." (*JPN*, p. 2.) ... "in conformity with the plain verdict of experience, the nature of reals is admitted to be made up of both the elements—universal and the particular and to be cognised as such by perceptual knowledge."...(*JPN*, p. 3.)

"Things are neither exclusively particulars, nor are they exclusively universals, but they are a concrete realization of both. The two elements can be distinguished by reflective thought, but cannot be rent asunder. So our experience of one particular individual is not confined to that individual alone, but extends to unperceived individuals also in so far as the latter typify the universal as a part of their constitution. Individuals, even when they belong to a class, will vary from one another. ...Repetition of experience only helps us to take stock of the universal in its true character, but once the latter is known, it does not stand in need of verification or confirmation by further observation".[2] (*JPN*, p. 6.)

4.3. The Jaina emphasizes the multiple nature of reality and accepts the standpoint of non-absolutism : "He asserts that neither unity nor diversity sums up the nature of a real, but both taken together do it. Unity is not exclusive of diversity or vice versa. The difficulty that is confronted is not grounded upon objective reality, but arises from a subjective aberration, which consists in the imagination of inconsistency between unity and diversity. But unity is associated with diversity and diversity is never found apart from unity, which is its very foundation." (*JPN*, p.58.)

"The central thesis of the Jaina is that there is not only diversity of reals, but each real is equally diversified. Diversification as induced by relations has been explained. The conclusion is legitimate that each real is possessed of an infinite number of modes at every moment. The number of reals is infinite and consequently their relations with one another are infinite. ... All things are related in one way or the other and ... relations induce relational qualities in the relata, which accordingly become infinitely diversified at each moment and throughout their career. ... Things are neither momentary[1] nor uniform[2]" ... (*JPN*, p.70.) According to the Jaina "a real changes every moment and at the same time continues. The continuity never breaks down." (*JPN*, p. 70.)

---

[1] *dravya* = substance; *kṣetra* = space ; *kāla* = time; *bhāva* = attribute.

[2] I may draw attention to the similarity of these ideas to the concept of an "individual element" in relation to the "population" in modern statistical theory.

"A real is that which not only originates, but is also liable to cease and at the same time capable of persisting. Existence, cessation, and persistence are the fundamental characteristics of all that is real. ... This concept of reality is the only one which can avoid the conclusion that the world of plurality, which is the world of experience, is an illusion." (*JPN*, p. 72.)

4.4. The relativism of the Jaina philosopher is to be sharply contrasted with some of the other Indian systems of philosophy.

"The Vedantist starts with the premise that reality is one universal existence; the Buddhist fluxist[3] believes in atomic particulars, each absolutely different from the rest and having nothing underlying them to bind them together. The Naiyāyika[4] believes both to be combined in an individual, though he maintains that the two characters are different and distinct. ... The Jaina differs from them all and maintains that the universal and the particular are only distinguishable traits in a real, which is at once identical with and different from both." (*JPN*. p. 13.)

It is, however, necessary to notice that :

"There is a difference—and intrinsic difference at that—between a manifested and an unmanifested real. ... They are identical and different both—identical in so far as it is the same substance and different in so far as it undergoes a change of characteristic. This is the Jaina position of non-absolutism." (*JPN*, p. 39.)

"A real is not entirely expressible in all its aspects and modes. But it is not inexpressible altogether. A real being a multiple entity is expressible and inexpressible both in reference to different aspects: it is expressible in so far as it partakes of a universal and is inexpressible so far as it is a unique individual."[5]...(*JPN*, p.113.)

"The unique individuality of a real is not accessible to conceptual thought and, hence, to language, but it is reached by an analysis of the nature of reality as it is apprehended in perception. ... we have tried to prove, following the guidance of the Jaina philosophers, that the nature of reals, on analysis, has been found to exhibit the following traits, viz., existence, non-existence and inexpressibility." (*JPN*, p. 127.)

---

[1] There is one well-known school of Buddhist philosophy which holds that reality consists of an infinite sequence of "atomistic" or completely independent "moments" which have no connexion with one another.

[2] On the other hand, the monistic philosophy of the Vedāntist holds that the Absolute transcends all change.

[3] The phrase "fluxist" requires a little explanation. The Buddhist school of philosophy (to which reference has been made in footnote) is known as *kṣaṇa-vāda* which means literally the theory or philosophy of "moments". It has been translated by S. Mookerjee as "fluxist" which, however, does not seem to be entirely happy.

[4] Another well-known school of Indian philosophy.

[5] It may be noted that a single or unique individual as such (that is without any relation to a "population" or "universe") has no meaning in modern statistical or probability theory.

5. RELATIONAL ASPECTS

5.1. Relational aspects have received special notice in Jaina logic.

"Everything is related with every other thing, and this relation involves the emergence of a relational quality. The qualities cannot be known *a priori*, though a good number of them can be deduced from certain fundamental characteristics." (*JPN*, p 3.) ..."A real is only a part of a system knitted together by a network of relations, from which it cannot be divorced." (*JPN*, p. 109.) "Every real is thus hedged round by a network of relations and attributes, which we propose to call its system or context or universe of discourse, which demarcates it from others." (*JPN*, p. 114.)

"It is idle to raise questions of chronological status as to whether the unity is prior to the elements or the elements are prior to the unity. In the concrete real at any rate they are coordinate. This unity of being and non-being, or rather of self-being and negation of other-being, is beyond the reach of logical concepts, and, hence, of linguistic symbols, which are but the vehicles of such concepts. The Jaina in recognition of this inalienable character of reals declares them to be inexpressible. The inexpressible may be called indefinite from the standpoint of formal logic. But this is not the whole character of a real. It is also expressible and logically definable as existent, as non-existent.[1]" (*JPN*, p. 115.)

"The Jaina conception of relation may be summed up as follows. Relations are objective verities which are as much given to intuition and to thought as the terms are. A relation has no objective status outside the terms. It is the result of an internal change in the nature of the terms. It is *sui generis* in that it cannot be placed under the head of identity or of difference, both of which are contained as traits in its being." (*JPN*, p. 211.)

5.2. The Jaina view of relatedness of the things is very naturally extended to the discussion of causality.

"...neither synchronism nor succession is believed by the Jaina to be the essential characteristic of causal relation. Causality is a relation of determination. The effect is that whose coming into being is necessarily determined by the being of another. The determinant is called the cause and the determinatum is called the effect. The determinant may be synchronous with the determined or may be separated by an interval ..." (*JPN*, p. 212.)

"What is the organ of the knowledge of causality? The Jaina answers that it is perception of the concomitance in agreement and difference. ... The Jaina takes the observation of concomitance in agreement and in difference to be one observation. ... The Jaina posits a twofold cause for the perception of universal relation—an internal and an external condition. The internal condition is found in the developed state of our mind and the external condition is the repeated observation of the sequence of the two events." (*JPN*, p. 217.)

---

[1] The Jaina view insists on the inadequacy of formal logic by introducing the concept of indefiniteness or indetermination or uncertainty as an inalienable character of reals; but also emphasizes the possibility of defining reals in terms of existent or non-existent taken together.

..."Such concepts as causality, substance, attribute and the like, are no doubt the ways in which the mind works up the data of experience, but this does not mean with the Jaina that they are true of the mind only and not of the extra-mental reality which they purport to understand. The Jaina would take them to be the instruments of discovery of the nature of reality, internal and external, which render the same kind of service as the sense-organs do." (*JPN*, p. 217.)

..."The different categories, viz., the selves, matter, time, space and so on, are deductions from experimental data. They have been posited since general concepts presuppose their existence and since without these principles the data of experience cannot be organized into a system. These categories in spite of their general and comprehensive character are not only not inconsistent with the existence of individual entities, but on the contrary they are entirely based on the objective data. Without the individual existents these categories would be reduced to unmeaning class concepts. The affirmation of categories as objective principles is thus proof of the existence of individual reals, which are included within the ambit of these categories. Without the individuals forming their contents the categories would be empty and barren, and the individuals without the categories would be reduced to a welter of chaos. The Jaina is a believer in plurality no doubt, but that plurality is not an unrelated chaos. The plurality is a system inasmuch as each individual is cemented with the rest by definite bonds of relationship." (*JPN*, pp.299-300.)

"From the analytic point of view (*paryāyārthikanaya*) the world is an infinite plurality with their infinite variations and modes. But the analytic view does not give us the whole nature of reality as it is. It is a partial picture that we derive of the world by means of such approach. The whole gamut of reality, however, reveals its universal unitive nature *as one existence* when it is envisaged from the synthetic angle of vision (*dravyārthikanaya*)." (*JPN*, p.301.)

"It seems legitimate to conclude that the universe is one existence which manifests itself, as substance (*dravya*) as it unifies the modes and attributes. The selfsame existence again reveals itself as Space in so far as it provides accommodation for the infinite plurality of existence within itself (*kṣetra*). It is the same existence which manifests itself as Time (w.f. *kāla*) is so far as it changes into aspects, as precedent and consequent as earlier and latter, as present, past and future modes. It is the same existence that evolves as phases and modes, attributes and states. The substance, time, space, attribute and relation are thus evolved from the same existence. The different categories, thus viewed as functional variations of one principle, are no longer in a position of antagonism of indifferent isolation. (*Aṣṭasāhasrī*, p. 113.)

The world of reals is thus not only plurality but a unity also. It is one universe that the Jaina metaphysics gives us. But the oneness is not secured at the sacrifice of the many, nor are the many left in unsocial indifference." (*JPN*, pp. 301-302.)

5.3. It has been observed that "Jaina philosophy is ... entitled to be called the paragon of realism. If experience be the ultimate source of knowledge of reality and its behaviour, we cannot repudiate the plurality of things. The admission of plurality necessitates the recognition of the dual nature of reals as constituted of 'being' and 'non-being' as fundamental elements. One real will be distinguished from another real and this distinction, unless it is dismissed as error of judgment, presupposes that each possesses a different identity, in other words that being of one is not the being of the other. This truth is propounded by the Jaina in that things are real, so far as they have a self-identity of their own unshared by others (*svarūpasattā*), and they are unreal in respect of a different self-identity (*pararūpasattā*). ... The logic of Jaina is empirical logic, which stands in irreconcilable opposition to pure logic.[1]" (*JPN*, p. 181.)

5.4. J. Sinha (*HIP*., vol. II, p.180) gives the following summary of Jaina philosphy: "The world is self-existent and eternal. All objects of the world are multiform (*anekānta*) and endued with infinite qualities and relations (*anantadharmaka*). This is relative pluralism. The reality can be considered from different points of views or *nayas*. The *nayas* are the standpoints. ... All judgments are relative and probable. No judgments are absolute. This is *syādvāda*. There are seven ways of predication. This is called *saptabhangīnaya*.[2]"

It is not strange that Jainas believe that "the different systems of philosophy are only partial views of reality. Jainism is the complete view of reality." (J. Sinha, *HIP*, vol., II, p. 180.)

6. Some general observations

6.1. I have given actual quotations from books on Jaina philosophy to convey the thoughts in their original form (of course, in English translation) without the bias of any subjective interpretations. I should now like to make some brief observations of my own on the connexion between Indian-Jaina views and the foundations of statistical theory. I have already pointed out that the fourth category of *syādvāda*, namely, *avaktavya* or the "indeterminate" is a synthesis of three earlier categories of (1) assertion ("it is"), (2) negation ("it is not"), and (3) assertion and negation in succession. The fourth category of *syādvāda*, therefore, seems to me to be in essence the qualitative (but *not* quantitative) aspect of the modern concept of probability. Used in a purely qualitative sense, the fourth category of predication in Jaina logic corresponds precisely to the meaning of probability which covers the possibility of (a) something existing, (b) something not-existing, and (c) sometimes existing and sometimes not-existing. The difference between Jaina "*avaktavya*" and "probability" lies in the fact that the latter (that is, the concept of probability) has definite quantitative implications, namely, the recognition of numerical frequencies of occurrence of (1) "it is", or of (2) "it is not"; and hence in the recognition of *relative* numerical frequencies of the first two categories (of "it is" and "it is not") in a synthetic form. It is the explicit recognition of (and emphasis on) the concept of numerical frequency ratios which distinguishes modern statistical theory from the Jaina theory of *syādvāda*. At the same time it is of interest to note that 1500 or 2500 years ago *syādvāda* seems to have given the logical background of statistical theory in a qualitative form.[1]

---

[1] Pure logic in the sense of formal logic.

[2] J. Sinha, *HIP*, vol. II, p. 180.

6.2. Secondly, I should like to draw attention to the Jaina view that "a real is a particular which possesses a generic attribute". This is very close to the concept of an individual in relation to the population to which it belongs. The Jaina view in fact denies the possibility of making any predication about a single and unique individual which would be also true in modern statistical theory.

6.3. The third point to be noted is the emphasis given in Jaina philosophy on the relatedness of things and on the multiform aspects of reals which appear to be similar (again in a purely qualitative sense) to the basic ideas underlying the concepts of association, correlation and concomitant variation in modern statistics.

6.4. The Jaina views of "existence, persistence, and cessation" as the fundamental characteristics of all that is real necessarily leads to a view of reality as something relatively permanent and yet relatively changing which has a flavour of statistical reasoning. "A real changes every moment and at the same time continues" is a view which is somewhat sympathetic to the underlying idea of stochastic processes.

6.5. Fifthly, a most important feature of Jaina logic is its insistence on the impossibility of absolutely certain predication and its emphasis on non-absolutist and relativist predication. In *syādvāda*, the qualification "*syāt*", that is, "may be" or "perhaps" must be attached to every predication without any exception. All predication, according to *syādvāda*, thus has a margin of uncertainty which is somewhat similar to the concept of "uncertain inference" in modern statistical theory. The Jaina view, however, is essentially qualitative in this matter (while the great characteristic of modern statistical theory is its insistence on the possibility and significance of determining the margin of uncertainty in a meaningful way). The rejection of absolutely certain predication naturally leads Jaina philosophy continually to emphasize the inadequacy of "pure" or "formal" logic, and hence to stress the need of making inferences on the basis of data supplied by experience.

6.6. I should also like to point out that the Jaina view of causality as "a relation of determination" based on the observation of "concomitance in agreement and in difference" has dual reference to an internal condition "in the developed state of our mind"

---

[1] I think it is also proper to note the occasional occurrence of certain intriguing phrases in the mediaeval period of Indian logic. As an example, I am giving below what I myself heard about 20 years ago from the late Dr. Sir Brajendra Nath Seal (the great Indian savant and the author of the "History of Positive Sciences of the Hindus" and other works who died in 1938). Dr. Seal told me that in a mediaeval Indian treatise there is a discussion about the practice of giving alms to Brahmins; and the question is raised whether the recipients of the gifts are always deserving persons. It is stated in reply that the practice of giving alms can be supported because "only ten out of hundred recipients are undeserving". I do not know whether the above phrase is to be interpreted as a simple statement about the number or proportion of recipients who were found to be undeserving or whether the phrase has any implications of a statistical or probabilistic nature. Dr. Seal's view was that the above phrase had some probabilistic significance but only in a latent or implied (but not explicit or developed) form. As the subject of mediaeval Indian logic is not my special field of study I have not had the opportunity of making more detailed enquiries. I am, however, mentioning this point because it may be worth while making some further researches in this matter.

(which would seem to correspond to the state of organized knowledge in any given context) and also to an external condition based on "the repeated observation of the sequence of the two events" which is suggestive of a statistical approach.

6.7. Finally, I should draw attention to the realist and pluralist views of Jaina philosophy and the continuing emphasis on the multiform and infinitely diversified aspects of reality which amounts to the acceptance of an "open" view of the universe with scope for unending change and discovery. For reasons explained above, it seems to me that the ancient Indian-Jaina philosophy has certain interesting resemblances to the probabilistic and statistical view of reality in modern times.

REFERENCES

Das Gupta, S. (1922): *A History of Indian Philosophy*, **1**, Cambridge University Press.

Mookerjee, Satkari (1944): *The Jaina Philosophy of Non-Absolutism*, Bhāratī Jaina Pariṣat, Calcutta.

Sinha, Jadunath (1952): *History of Indian Philosophy*, Central Book Agency, Calcutta.

Vidyabhusana, Satis Chandra (1921): *A History of Indian Logic*, Calcutta University.

——— (1909): *Nyāyāvatāra*, Indian Research Society, Calcutta.

THE SYĀDVĀDA SYSTEM OF PREDICATION

By J. B. S. HALDANE

University College, London

The search for truth by the scientific method does not lead to complete certainty. Still less does it lead to complete uncertainty. Hence any logical system which allows of conclusions intermediate between certainty and uncertainty should interest scientists. The earliest such system known to me is the *Syādvāda* system of the Jaina philosopher Bhadrabahu (?433-357 B.C.). Mahalanobis (1954) has commented on it. A central feature of this system is the *saptabhangīnaya* or list of seven types of predication. These are as follows.

(1) *syādasti.*	May be it is.
(2) *syātnāsti.*	May be it is not.
(3) *syādasti nāsti ca.*	May be it is and is not.
(4) *syādavaktavyah.*	May be it is indeterminate.
(5) *syādasti ca avaktavyaśca.*	May be it is and is indeterminate.
(6) *syātnāsti ca avaktavyaśca.*	May be it is not and is indeterminate.
(7) *syādasti nasti ca avaktavyaśca.*	May be it is, is not, and is indeterminate.

Mahalanobis illustrated this from the throw of a coin, and held that it could serve as a foundation for statistics. However I wish to show that it arises naturally in simpler cases, including simple cases where the affirmative predication *asti* would be "This is hot", or "This is a man".

In any such case an uncertain judgement is usually somewhat quantitative, as in "I think this is a man, though it may be a statue." I therefore begin with a very abstract field, that of algebra. Here we may be certain of our answer. If $x+2 = 3$, then $x = 1$. But if $x^2-3x+2=0$, then $x = 1$ or 2. We cannot say that the probability that $x = 1$ is greater than, less than, or equal to the probability that $x = 2$. Further data may lead to either of these judgements. Five hundred years ago one might perhaps have spoken of indeterminate solutions of equations. Thus if $x^3-x^2+x-1 = 0$, $x = 1$ or $\pm\sqrt{-1}$. The last two solutions were *avakta* (incapable of being spoken) until the invention of complex numbers. Today we can find better examples in the field of finite arithmetic.

Consider the finite arithmetic *modulo* m. The only admissible values of a variable are the m residues $0, 1, 2, \dots m-2, m-1$, that is to say the possible remainders after division by m. For example *modulo* 5, $4+3 = 7 = 5+2$, so we write $4+3 \equiv 2$. And $4\times3 = 12 = 2\times5+2$, so we write $4\times3 \equiv 2$. Let us consider the theory of functions *modulo* m. We can define any function $f(x)$ by a table of the values which it assumes for the different admissible values of x. Thus the function $3x$ (*mod.*5) can be defined by the table 0, 3, 1, 4, 2. For example if $x = 4$, $3x = 2$. Of course many other functions are identical with it. For example $3x^5 \equiv x^9+2x^5 \equiv 3x$. A function which assumes all the admissible values unequivocally is called biunivocal[1], and it is easy to show that there are $m!$ biunivocal functions. However some functions are univocal, but their inverses are not. In this case some residues do not occur in the table, while others occur more than once. For example the table of $3x^2+1$ (*mod.*5) is 1, 4, 3, 3, 4. The number of univocal functions is m^m, since each place in the table can be filled in m ways.

If a function is not univocal, but its inverse is univocal, we obtain a table such as that for $x^{\frac{1}{2}}$, namely 0, 1 or 4, अ, अ, 1 or 4. Here I introduce the symbol अ, for *avakta*, for an undefined number. There is no number whose square (*mod.* 5) is 2 or 3. अ may occur in a table as an alternative to a number. For example the function e^x is never integral when x is a residue other than zero. Nor is it integral for most values of x which are *avakta*, such as $\sqrt{2}$. But it is integral for such numbers as log 2. Hence the table of e^x is 1; अ; अ; अ; अ; and 0, 1, 2, 3, 4, or अ. The last place in the table corresponds to $x =$ अ. Similarly if we consider the function y defined by $y^3-y^2 = x^2$, then we find the table

0 or 1, अ, 2 or अ, 2 or अ, अ, 3 or 4 or अ.

For when $x = 2$ or 3, $y^3-y^2-4 \equiv 0$, so $y = 2$ or $2\pm\sqrt{2}$, the latter two roots being congruent with $\frac{1}{2}(-1\pm\sqrt{-7})$. These quantities are inexpressible (*avakta*) *modulo* 5. And when $x =$ अ, $x^2 = 2, 3,$ or अ, so y may be 3 or 4, as well as अ.

Thus for a full enumeration of functions *modulo m* we need a table with $m+1$ places corresponding to the residues $0, 1, 2 \ldots, m-1$, and अ. In each place we can set one, or any number, of these symbols, but we must set at least one. So each place can be filled in $2^{m+1}-1$ ways, for each of the $m+1$ symbols can be present or absent, except that all cannot be absent. Thus the total number of functions *modulo m* is $(2^{m+1}-1)^{m+1}$, for example 62, 523, 502, 209 if $m = 5$, as compared with only 120 biunivocal functions, and 3125 univocal.

Now consider the simplest of the finite arithmetics, namely arithmetic *modulo* 2. There are only two elements, 0 and 1. Electronic calculators are based on this arithmetic. These machines are so designed that each unit, as the result of any instruction, will be active (1) or inactive (0) at any given moment. And it is possible, in principle, to predict whether it will be active or inactive. That is to say ambiguity is avoided, and the machine is designed to operate in terms of univocal functions. Nevertheless it is possible to provide such a machine with an instruction to which it cannot give a definite answer. It is said that some such machines, when given an instruction equivalent to one of the paradoxes of Principia Mathematica, come to no conclusion, but print 101010...indefinitely. Clearly a machine could be designed to print अ in such a case. It is obviously possible to design a machine which would print "0 or 1" in response to the instruction $x^2-x = 0$. A machine with the further refinement suggested above would respond "0, 1, or अ" to the instruction "$(x^2-x)\cos x \equiv 0 \pmod 2$". Such a machine could give any of 7 responses, namely :

0, 1, अ, 0 or 1, 0 or अ, 1 or अ, 0 or 1 or अ.

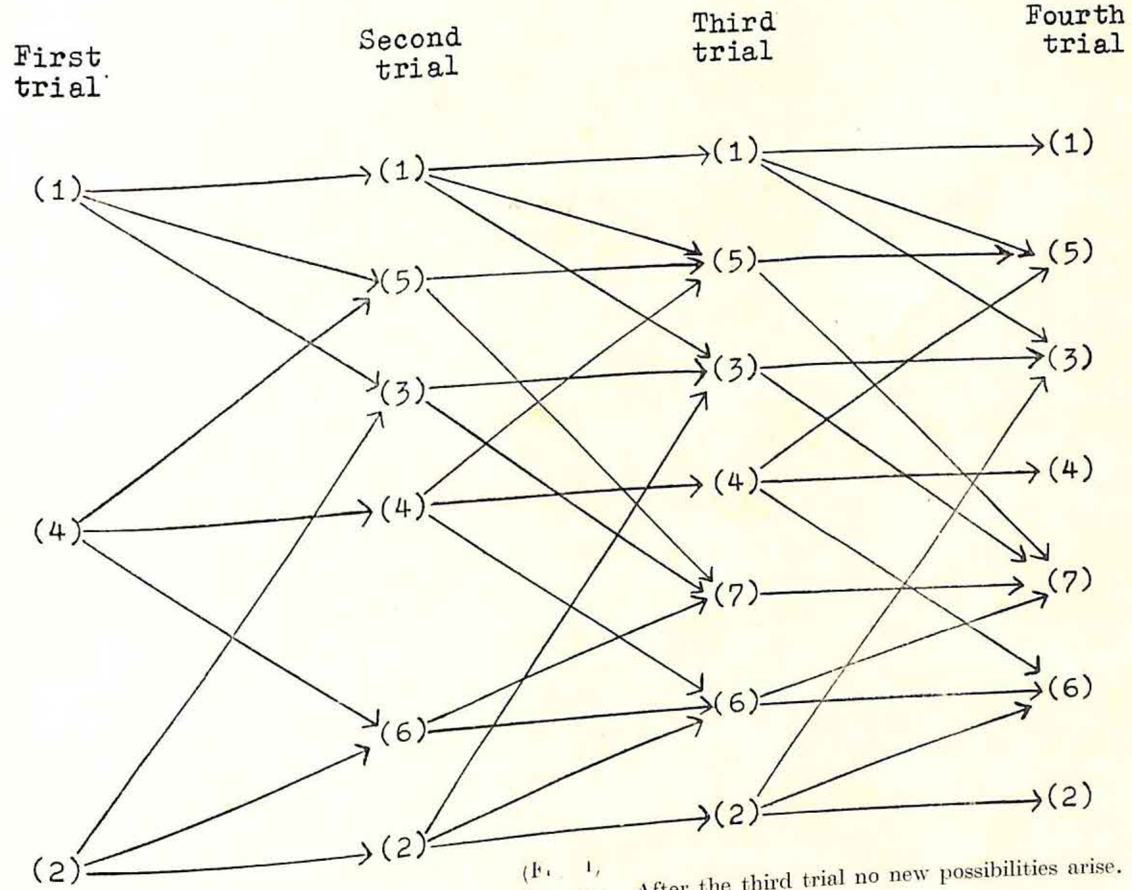
These are the *saptabhanghīnaya* with the ommission of the syllable *syad*.

---

[1] I have deliberately chosen a word with the same root (Latin *vox* = Sanskrit *vak*) as *avakta*.

I now pass to an example where the *saptabhanghīnaya* is actually applied in scientific research, and which I suspect is not far from what was in Bhadrabahu's mind. In the study of the physiology of the sense organs it is important to determine a threshold. For example a light cannot be seen below a certain intensity, or a solution of a substance which is tasted as bitter when concentrated cannot be distinguished from water when it is diluted. Some experimenters order their subjects to answer "yes" or "no" to the question "Is this illuminated ?", or "Is this bitter ? ". If the experimenter is interested in the psychology of perception he will permit the subject also to answer "It is uncertain", or some equivalent phrase. The objection to this is that some subjects may do so over a wide range of intensities.

Now consider a subject who is shown a series of illuminated patches, some above his threshold of perception, some below it, and others very close to it, in a randomised series. We will suppose that he is in a steady state of sensory adaptation, that he replies in Sanskrit, and that he is aware that his answers will sometimes be incorrect. At any given trial he will answer "*syādasti*", "*syādavaktavyaḥ*", or "*syātnāsti*". After the second trial of a light of an intensity near the threshold he may have given two of these answers, for example "*syāt-nāsti ca avaktavyaśca*". After the third he may have given all three, though this is not very probable. The possibilities may be schematised as follows:



(Fig. 1)

Diagram showing the possible sequences of predications. After the third trial no new possibilities arise.

It is clear that the seven possibilities are exhaustive. My only possible criticism of Bhadrabahu is this. If the subject has given two different answers he may be aware that there is a possibility that he will later give the third. Once he has given all three, no further possibility is open. It might therefore be argued that the seventh type of predication should be "*asti nāsti ca avaktavyaśca*".

On the hypothesis that the subject is in a steady physiological and psychological state, the probabilities of each of the three answers to any given stimulus are constant. Let the probabilities of answering *syādasti*, *syādavaktavyah* and *syātnāsti* be p, q, and r, where $p+q+r = 1$. If, after n trials, the probabilities of the 7 types of predication are $P_{1,n}$, $P_{2,n}$, etc. where $P_{1,n}$ is the probability of *syādasti*, etc., then the vector $[P_{1,n};\ P_{2,n};\ P_{3,n};\ P_{4,n};\ P_{5,n};\ P_{6,n};\ P_{7,n}]$ is transformed into the vector $[P_{1,n+1};\ P_{2,n+1};\ \ldots$ etc.$]$ by multiplication by the matrix

$$\begin{pmatrix} p & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & r & 0 & 0 & 0 & 0 & 0 \\ r & p & p+r & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q & 0 & 0 & 0 \\ q & 0 & 0 & p & p+q & 0 & 0 \\ 0 & q & 0 & r & 0 & q+r & 0 \\ 0 & 0 & q & 0 & r & p & 1 \end{pmatrix}$$

Evidently this could be made a little more symmetrical by transposing row and column (3) and (4).

The latent roots of this matrix are :

$$1,\ p+q,\ q+r,\ p+r,\ p,\ q,\ r.$$

So

$$\begin{aligned} P_{1n} &= p^n \\ P_{2n} &= r^n \\ P_{3n} &= (p+r)^n - p^n - r^n \\ P_{4n} &= q^n \\ P_{5n} &= (p+q)^n - p^n - q^n \\ P_{6n} &= (q+r)^n - q^n - r^n \\ P_{7n} &= 1 - (q+r)^n - (r+p)^n - (q+r)^n + p^n + q^n + r^n. \end{aligned}$$

Thus unless one of $p, q,$ or r is zero the final predication will be *syādasti nāsti ca avaktavyaśca*. In many cases when the stimulus is far from the threshold, p or r will be unity. The subject will always, or never, say "this is bitter", or "this is illuminated". It is unlikely that q will ever be unity. So in this case *syādavaktavyah* will almost always be at best a provisional predication. It is however possible that p or r (say r) should be small, but not zero. If so $P_{5,n}$ will reach a maximum for some value of n and then decline. For example if $p = .6$, $q = .3$, $r = .1$, P_5 reaches its maximum value of .5184 when $n = 4$.

I have dealt with a case which arises when the question asked is as simple as possible. Human judgements are generally more complicated. We may attend to the data of several different senses, and of our memories. Thus we arrive at one conclusion from one set of data, and another from another set. We say that wood is hard when compared with clay, soft when compared with iron, indeterminate when compared with similar wood.

The close analysis of vision with a dark adapted eye shows that in this case at least, Mahalanobis was correct in regarding the *saptabhangīnaya* as foreshadowing modern statistical theory. It appears that when dark adaptation is complete, about five quanta of radiation must arrive within a short time in a small area of the retina before light is reported. Whether they will do so with a given intensity of illumination can only be stated as a probability. It is probable, though not by any means certain, that more complicated judgements depend on similar probabilities of events within the central nervous system.

Whatever philosophers of other schools may think, a Jaina can hardly object to regarding human predication as a special kind of animal behaviour. In this he agrees with followers of Darwin, such as myself. Attempts at a logical classification of animal behaviours frequently lead to a separation of 2^n-1 types, where however n may exceed 3. Thus Haldane (1953) classified the possible results of learning in an animal as follows. In any situation an animal will, or will not, give a certain response R, say eating a particular type of food within a minute of its presentation, or lifting its leg within ten seconds after an auditory signal is given.

If we compare the set of possible situations in which an animal may be placed before and after an experience E, they fall into four categories, $r\,r$, rR, $R\,r$, and RR. A situation $r\,r$ is one in which the response is not given before or after the experience. A situation $r\,R$ is one which it is given after E, but not before E, and so on. All situations may be $r\,r$. For example no-one has taught a dog to write. Some may be $r\,r$ and some rR. For example a dog which did not previously bring objects from the water to his master can learn to do so on command. In Pavlov's experiments a dog which previously only salivated (R) when given food will do so when certain auditory or other stimuli are given. Thus for such a dog all situations fall into the classes $r\,r$, RR, and rR. It can easily be seen that the effect of any experience on an animal can be classified according as the situations in which it can be placed fall into one, two, three, or all four of these classes. There are thus 2^4-1, or 15 qualitatively different results of an experiment in which an attempt is made to alter an animal's behaviours. In this classification the animal is assumed never to give an indeterminate response. If it can do so, both before and after, there are, as I pointed out, 2^9-1, or 511 possible results. The same principles may be applied to the comparison of the behaviour of two different animals, or two different races or species.

It is foolish to pretend that ancient philosophers anticipated all modern intellectual developments. And I believe that we, today, can do more honour to their memories by thinking for ourselves, as they did, than by devoting our lives to commentaries on them. But if we do so it is our duty to point out cases where it turns out that our own thought has run parallel to theirs. I was unaware of Bhadrabahu's existence when I wrote the paper to which I refer. The fact that I reached a conclusion so like his own suggests that we may both have seen the same facet of many-splendoured truth.

No doubt we reached it by very different methods, Bhadrabahu by meditation, I by thinking about the results of concrete experiments on animals. Such methods will often lead to different conclusions. This was the view of Warren Hastings in his introduction to Wilkins' translation of the Bhagavad Gita.

"But if we are told that there have been men who were successively, for ages past, in the daily habit of abstracted contemplation, begun in the earliest period of youth, and continued in many to the maturity of age, each adding some portion of knowledge to the store accumulated by his predecessors, it is not assuming too much to conclude, that as the mind ever gathers strength, like the body, with exercise, so in such exercise it may in each have acquired the faculty to which they aspired, and that their collective studies have led them to the discovery of new tracks and combinations of sentiment, totally different from the doctrines with which the learned of other nations are acquainted: doctrines, which however speculative and subtle, still, as they possess the advantage of being derived from a source so free from every adventitious mixture, may be equally founded on truth with the most simple of our own".

If, on the other hand, the contemplation of one's own mind, and that of the minds of animals, lead to similar results, such results are, perhaps, worthy of serious consideration.

References

Mahalanobis, P. C. (1954): The foundations of statistics. *Dialectica*, **8**, 95-111.

Haldane, J. B. S. (1954): A logical analysis of learning, conditioning, and related processes. *Behaviour*, **6**, 256-270.

Paper received : November, 1956.

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Edited by : P. C. MAHALANOBIS

VOL. 18, PARTS 3 & 4 SEPTEMBER 1957

THE UNDERWORLD OF PROBABILITY

By SIR RONALD A. FISHER

University of Cambridge

Readers of my recent book *Statistical Methods and Scientific Inference*, may have noticed that in the author's opinion, stagnation and confusion of thought over a long period have been due to the common failure to distinguish various different types of evidence, and the correspondingly various types of uncertainty appropriate to an appreciation of such evidence.

The following paper is much less than an exploration of the varieties of logical uncertainty in normal experience. It takes one short step towards a recognition of these varieties and introduces the beginnings of a classification, more detailed than that of my book. The illustrations given will, I believe, help others to engage in a deeper penetration.

1. A RANKING OF THE LEVELS OF UNCERTAINTY

A series of meaningful mathematical statements may be made, each differing from the one before by its greater uncertainty, as follows :

Certainty. The mean distance of the Sun is x miles.

Uncertainty of Rank A. The mean distance of the Sun is known as a Random Variable. For all x we have a known function $P(x)$ giving the probability that the distance is less than x. If $P(x)$ is continuous there will thus be a known function x_P, such that the probability that the true distance is less than x_P shall be P. All percentile points of the distribution of the random variable are known with exactitude. The symbolical statement

$$Pr(x < x_P) = P,$$

is then justified by the known functional relationship between P and x_P.

Uncertainty of Rank B. The percentile points of the distribution spoken of in A above, are not known, but probability statements may be made about them. Each then is a random variable, so that $x_{PP'}$ is the value of x such that

$$Pr(x_P < x_{PP'}) = P',$$

or, in a more elaborate notation,

$$Pr\{Pr(x < x_{PP'}) > P\} = P',$$

and $x_{PP'}$ is a known function of P and P'.

Uncertainty of Rank C. If, however, $x_{PP'}$ is not known with certainty in terms of P and P', but if given P and P' it has a known distribution as a random variable, the uncertainty may be said to be of rank C.

Evidently an unending series of such ranks is defined by an extension of the same process.

Since in statistical discussion it is common to speak of a *probability* as an unknown quantity, appropriate ranks should be assigned to such statements as follows :

Certainty. The Event will occur.

Uncertainty of Rank A. The Event has a known probability, π.

Uncertainty of Rank B. For all P a function π_P is known such that the probability that π is less than π_P is exactly P.

Uncertainty of Rank C. π_P is given as a random variable, such that a known function of P and P', namely $\pi_{PP'}$ shall exceed π_P with frequency P'.

In respect to the success or failure of an Event, it will be noticed that the statements of probability *a priori* and *a posteriori* of Bayes are both statements with uncertainty of Rank B. A gambler's statement that he has a probability of one in 36 of throwing a pair of sixes leaves the Event with the simplest kind of uncertainty, of Rank A.

2. BETS AND SIDE-BETS BASED ON UNCERTAIN INFORMATION

It might be thought that information at the levels of deeper uncertainty, C, D, E etc., is necessarily so tenuous and remote that there is no practical advantage to be gained by possessing it. I do not believe that such an opinion should be formed hastily. The gain immediately in view is, indeed, no more than the clarification of thought, but such clarification is not necessarily without practical advantages. The primitive method of exploiting a knowledge of probability (Rank A) for personal advantage is by laying advantageous bets; for although every particular event is recognized to be uncertain, a man (A) who believes that, in throwing a die, an Ace will be turned exactly once in six trials, will *ipso facto* be aware that if he can bet four to one against the Ace, or 4 to 21 in its favour, though he may often lose, yet on his data, and in the long run, his bet will certainly be an advantageous one.

Now it may be that the information available is not so certain as *A* supposes; a second gambler *B* may be aware that there is variability among the dice in the box from which *A* has chosen one at random, and although he may have no knowledge of which has been chosen, if he has, or thinks he has, information to the effect that quite 10% of the dice in this box will throw an ace more than once in five trials, he will feel justified in accepting odds at 9 : 1 or higher in betting that *A* will lose in the long run, if he continues with the same die in his present course of action.

It is easy then to perceive that *C*, who does not know and does not think he knows the percentile points of the probability of throwing an Ace, but believes he knows how the relevant percentile point will be distributed in a consignment of boxes, one of which has been used to supply the die in question, may know with confidence what odds he can profitably accept in betting against the success of *B*'s wager. Nor is *C*'s information of the lowest rank that could be profitably utilized in such a sequence of side-bets.

Although the betting operations I have described cannot be exhibited, I believe, in any orthodox gaming establishment, for the reason chiefly that they take too long, yet in Commerce and Finance very similar operations are customary. A firm of contractors building a bridge will employ engineers who use tables of the strength of materials, and perhaps, empirical data on the frequency of destructive floods, hurricanes or earthquakes, of a strength capable of jeopardising their work. In view of recognised uncertainties of the strength of the structure, and its parts, they build in what are called "factors of safety" at great expense, to cover, as they hope, all reasonable risks. Nevertheless, the corporation operating such a structure will usually seek out an Insurance Office, who at an appropriate premium, or in other words at chosen odds, will underwrite the residual risk of the accidental destruction of the structure. They are in effect betting that for an assigned period the precautions of the engineers will be shown to have been sufficient.

Moreover, even the simple investor risking his savings by buying ordinary stock, perhaps in an Insurance Company, is in effect laying down a stake on the proposition that the company will collect enough in premiums to cover the large risks it has underwritten, and above this to pay a reasonable dividend on their ordinary shares.

The mathematical underworld I have introduced has a qualitative resemblance to the specification of risk in quite familiar operations in our own world. It is only the mathematical tradition that has been too shy of uncertainty, too firmly attached to the ethereal regions of mathematical certainty, to have attempted any very deep exploration.

It is important to recognize that the reference to a consignment of boxes as capable of justifying statements of uncertainty of rank *C* could in one respect easily be misunderstood, for an examination of a single box of dice could yield a probability integral such as the ogive below :

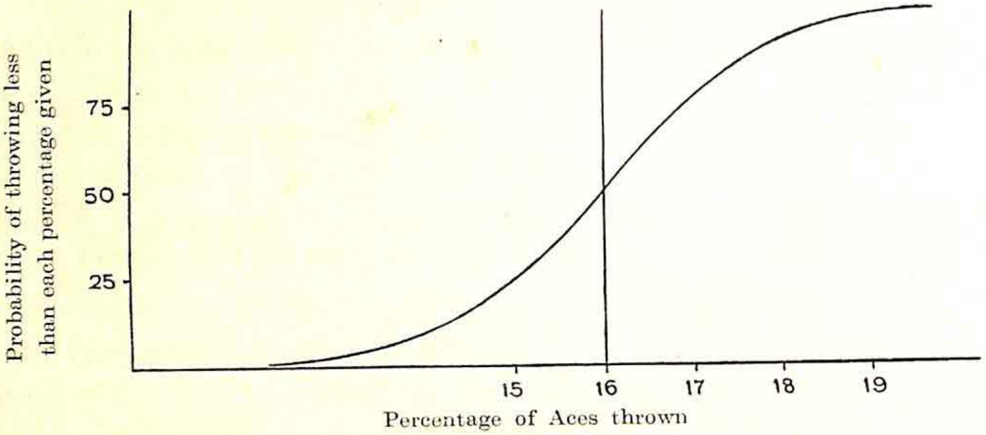


Fig. 1. Representing the kind of information on which B stakes his wager.

so that a Consignment of such boxes might be thought of as yielding of a sheaf of such ogives. This, however, is an unintended effect of an imperfect example. For uncertainty of rank C it is intended that there shall be variation, and well specified variation, in the probability integral as shown on the ordinate at 16% on the diagram, and likewise variation in the probability integral at other percentages, but it is not intended that C should possess knowledge of the covariation, or simultaneous variation, of these several ordinates, such as a sheaf of distinct ogives would imply.

Had the data supplied such a sheaf of ogives as would a consignment of boxes taken literally, the mean ordinate at each percentage would give a single ogive appropriate to the result of emptying all the boxes of a consignment into a single receptacle, and choosing a die at random from this. The uncertainty would then be of rank B. It might enable a gambler to bet successfully *against* B, and not merely *about* the success of B's wager. So long as only a single percentage of Aces is in discussion, and this percentage subject to betting at specified odds, this stipulation makes no difference. When inductive inferences have to be drawn, however, all possibilities must be comprehended.

3. THE PROBLEM OF BINOMIAL DATA

3.1. *General formulation*

Towards the end of the seventeenth century James Bernoulli considered at great length, in his posthumously published work *Ars conjectandi*, what inferences could properly be drawn from a finite empirical record of successes and failures, on the value of the true probability of success. Either the work was never completed, or the author encountered unexpected difficulties, for the promise of the early chapters was never fulfilled. Two generations later Thomas Bayes, by introducing, in addition to the data, the notion of probabilities *a priori* (regarded as possibly axiomatic) was able to deduce a definite frequency distribution for the probability, namely, for a successes and b failures out of n trials,

$$\frac{(a+b+1)\,!}{a\,!\;b\,!}\,p^a(1-p)^b dp\;;$$

so that all percentile points were exactly definable by the incomplete Eulerian integrals.

There are, indeed, cases, such as that laid down by Bayes for demonstrating his theorem, in which information of an *a priori* type based on a previous and independent act of random sampling, are really available; and for these cases Bayes' argument, and his conclusion, with uncertainty of rank B, are exact. In most cases in the Natural Sciences, however, such knowledge *a priori* is lacking, and all efforts to express the inference proper to a rational mind, derivable from the observations only, in terms of probability statements about the unknown probability p, have been unsuccessful. I propose to show that the exact inferences properly supplied by such data can be expressed in uncertainty statements of Rank C.

In finding what probability statements of whatever rank may be inferred about the unknown probability, π, we shall be concerned with the probabilities, in random samples, of all possible conjunctions of possible values of π, with all possible observational records. The actual observations recorded are then inserted as particular cases known to be of the general class of random samples, to provide the particular set of probability statements appropriate to the particular unknown from which they were derived.

For binomial data, showing a successes out of n trials, let c_r stand for term of the binomial expansion

$$\frac{n\,!}{r\,!\,(n-r)\,!}\,\pi^r(1-\pi)^{n-r}$$

and let

$$P = c_0+c_1+\,\dots\,+c_{a-1}+\lambda c_a.$$

P is then an explicit function determined by the observational data, by the unknown π, and by an auxiliary quantity λ which is to be distributed, independently of the observations, in a rectangular distribution from 0 to 1, so that independently of all else

$$Pr(\lambda < \lambda_1) = \lambda_1 \qquad 0 \leqslant \lambda,\, \lambda_1 \leqslant 1.$$

Simultaneous values of

(i) the observations a, b, $(a+b = n)$,

(ii) the unknown probability, π,

(iii) the random variable, λ,

constitute a contingency, and of such contingencies the probability that

$$P \equiv \sum_0^{a-1} c_r+\lambda c_a,$$

a known function of (π, λ) given the observations, is less than P' is simply equal to P'.

Bureau Ednl. Psy. Research
TRAINING COLLEGE
Dated ....
Acces. No ....

If P' were, for example, 1% then one per cent of all contingencies will have P less than P', and may be termed the first percentile of all contingencies. The boundary of this region, in which $P = P'$, defines the 1% value of π for each possible λ, namely $\pi_{P\lambda}$.

Since λ has a known distribution independently of all else, its variation defines the sampling distribution of $\pi_{P\lambda}$ as a random variable.

3.2. *Formal analysis*

In a more formal notation we may denote by $\pi_{P\lambda}$ that function of P and λ for which the known function of π and λ

$$\sum_{0}^{a-1} c_r + \lambda c_a$$

shall be equal to P, where $\pi_{P\lambda}$ is substituted for π. Then $\pi_{P\lambda}$ will increase monotonically with P, and decrease monotonically as λ is increased, for all values of P and λ, so that if P is less than P_1 it follows that

$$Pr(\pi < \pi_{P_1\lambda_1}) = P_1,$$

for $\lambda = \lambda_1$, and

$$Pr(\pi < \pi_{P_1\lambda_1}) > P_1$$

for $\lambda < \lambda_1$.

But, unconditionally

$$Pr(\lambda < \lambda_1) = \lambda_1.$$

Hence, dropping suffixes,

$$Pr\{Pr(\pi < \pi_{P\lambda}) > P\} = \lambda,$$

for all values of P and λ.

This is an uncertain, though mathematically precise, statement about π of Rank B, and therefore of Rank C, about the Event having probability π.

3.3. *Verification*

A concrete notion of the probability statements inferred from simple binomial data may be gained from the experimental process of verification. Such processes are always laborious and their specification is no more than a means of making clear the exact meaning of the inferences to one who wishes to understand their relation to a possible experimental test.

First, a random sample of n is taken and the number of successes counted. Secondly, a value of λ is chosen at random so that it shall fall with equal probability within all equal ranges from 0 to 1.

Next $\pi_{P\lambda}$ is calculated with P equal, for example, to .99, and B bets 99 to 1 that π, the true probability of success, is less than $\pi_{P\lambda}$. Comparing B's calculated value with the true value π, it is ascertained whether or not he has won his bet.

If for every possible sample he can encounter B gives to λ all possible values from 0 to 1 with even frequency, then it is demonstrable, and will be verified experimentally that B loses his bet with a frequency of just 1%. $\pi_{P\lambda}$ may properly be called the upper 1% value of π deducible from the observations. It is, however, not known save as a random variable.

Gambler C may now observe that, in the particular circumstances in which B is betting $\pi_{P\lambda}$ has a definite distribution determined by the known frequency distribution of λ. He has the material for betting with confidence on the 1% value found by B for any particular sample.

If it were thought that the values $\pi_{P\lambda}$, for the same value of λ for different possible samples, could be associated together as a recognizable sub-set of possibilities, this would associate the percentile points for different values of P so as to specify their covariation in a way which would imply uncertainty of no more than rank B. No probability statements, however, true for all π, can be deduced for such sub-sets.

3.4. *A numerical example*

As an example we may take a case such as that in which an event has happened three times only among 14 random tests. The possible upper one per cent points ($P = .99$) must lie in the range bounded by the value of p at which the first three terms

$$q^{14}+14pq^{13}+91\ p^2q^{12}$$

constitute exactly 1% of the distribution, and that for which the inclusion of the fourth term

$$364p^3q^{11}$$

brings the total to exactly this percentage. The value of p required in the first case is

$$p_1 \quad .478244,$$

and in the second case

$$p_2 \quad .556655.$$

These constitute the limits of the distribution; they provide the absolute inequalities satisfied by the upper 1% point of the possible values of the probability in the population from which the sample was drawn.

The simplest method of arriving at values of p (or q) which make the partial sum of a binomial series equal to an assigned percentage of the whole consists in the use of Table V of *Statistical Tables* as explained in the *Introduction* Ex. 2.

In the first case three out of the 15 terms of the binomial expansion are to be contrasted with the remaining twelve terms. Doubling these numbers we use 6 for n_1 and 24 for n_2. The table for the Variance Ratio at the 1% point then gives 3.67. Multiplied by 6/24 this corresponds to the ratio of "sums of squares" 0.9175, and so to $p = .9175/1.9175$ or

$$p_1 \quad .47849.$$

A more accurate value is obtained from the corresponding value of z given on the page facing, namely .6496; doubling this and taking the Naperian antilog, the variance ratio comes to 3.66637, and the lower limit of the distribution of the upper 1% point, to

$$p_1 \quad .478240$$

correct to two more places. The table of z, for the range of percentile values for which it is available, will be suitable for determining the end points of such distributions.

For chosen values of p between these end points, the corresponding values of λ will be required. These are calculated directly from

$$\lambda = (0.01 - c_0 - c_1 - c_2)/c_3$$

in which c_0, c_1, etc., stand for the successive binomial terms.

In the present case I find :

p%	λ
48	.01101
49	.07966
50	.15890
51	.25156
52	.36121
53	.49258
54	.65160
55	.84584

The values of λ supply the probabilities of the random variable falling below the values of p indicated. The ordinates of the curve are to be found from the value of

$$y = d\lambda/dp,$$

for constant values of P, and these in the case of the binomial distribution take the simple form

$$y = \frac{3}{p}(1-\lambda) + \frac{11}{q}\lambda,$$

in which the coefficients are the observed numbers of occurrence and non-occurrence of the primary event. The ordinates take the finite values.

$$\frac{3}{p_1} \text{ and } \frac{11}{1-p_2};$$

at the termini where p takes the values p_1 and p_2. The range must, generally speaking, be commensurate with the reciprocal of the total number observed. This is an order of magnitude which, as the sample size increases, tends to zero compared with the standard deviation of the estimate

$$\sqrt{\pi(1-\pi)/n}\,;$$

the recognition of uncertainty of Rank *C* may therefore be regarded as a refinement appropriate to the exact treatment of such samples.

Fig. 2 shows the distribution of the upper 1% point in the case considered.

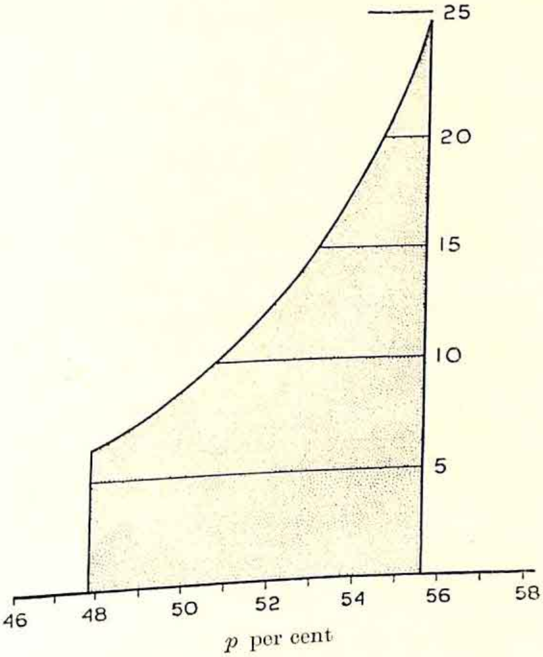


Fig. 2. Distribution of the upper 1% point indicated by the data for an unknown probability.
(lower limit 47.82, upper limit 55.67)

π is the true value of the probability of an event of which all that is known is that it has been observed 3 times in 14 trials.

(a) No value for π can be asserted.

(b) No value can be stated such that the probability of π exceeding that value is 1% (or any other percentage).

(c) Such a 1% value can, however, be specified as a random variable, as shown in this Figure.

It would not be legitimate to sample the distribution (*c*) provided by the observations with the intention of asserting that the particular value found was the 1% point provided by them for the distribution of π; it would be equally illegitimate, and for exactly the same reason, in cases such as that of Bayes, in which the data provide a distribution (*b*) of this unknown, to choose a value at random from this distribution, and to assert that the particular value found in this way was the value of π provided by the data; or, finally, if the true value of π were known, but the happening of the event were still uncertain, it would do nothing to diminish this uncertainty to use the apparatus of gambling to see whether an event with exactly this probability did or did not occur in a random trial. Such so-called randomisation tests, proposed in the name of Decision Functions, serve only to obscure the elucidation of the particular nature of the uncertainty in which observations of a given kind leave the rational observer.

In binomial sampling the case in which all the experimental tests give like results, as, for example, if no success had been recorded in 14 trials, deserves special notice. For in such cases it would be possible for π to be zero, so that the range of any percentile point is from zero to

$$p = 1-P^{1/14}$$

as its upper terminus.

Since now,

$$\lambda q^{14} = P,$$

it is clear that λ cannot be less than P, and that in a fraction P of all cases the percentile of the distribution of π represented by π_P is included in a condensation at zero. The remainder of the distribution may be represented as a frequency curve with area $1-P$, between zero and its finite upper bound.

The ordinate,

$$y = \frac{14}{q}\lambda$$

varies from

$$y = 14P \text{ to } 14/P^{1/14}.$$

REFERENCES

BAYES, T. (1763): An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, **53**, 370.

BERNOULLI, J. (1713): *Ars conjectandi*, opus posthumum, Basileae.

FISHER, R. A. (1956): *Statistical Method and Scientific Inference*, Oliver and Boyd, Ltd., Edinburgh.

FISHER R. A. AND YATES, F. (1953): *Statistical Tables for Biological, Agricultural and Medical Research*, 4th edition. Oliver and Boyd, Ltd., Edinburgh.

Paper received : September, 1957.

ON UNBIASED ESTIMATES OF UNIFORMLY MINIMUM VARIANCE

By R. R. BAHADUR

Indian Statistical Institute, Calcutta

1. SUMMARY

In a given statistical framework let T be the class of all estimates that are the uniformly minimum variance estimates of their respective expected values. Let T_b denote the class of bounded estimates in T. The main conclusions of the paper may then be outlined as follows. (i) There exists a statistic such that T_b is the class of all bounded functions of this statistic; moreover, every real valued function of this statistic is in T. It follows, in particular, that if t is in T_b and u is a real valued function of t, then u is in T. (ii) T contains an unbiased estimate of every estimable parameter if and only if the framework admits a complete sufficient statistic. In this case, as is well known, T is the class of all real valued functions of the complete sufficient statistic.

Conclusion (ii) can also be stated as follows. Suppose that in the given framework the maximum possible reduction of the sample space by means of sufficient statistics has already been carried out. Then either each estimable parameter has a unique unbiased estimate, or there exist estimable parameters that do not admit unbiased estimates of uniformly minimum variance.

A more precise statement and discussion of the above conclusions is deferred to later sections. The conclusions are established under the mild restriction that the sample space is, or may be taken to be, a subset of the m dimensional Cartesian space ($1 \leqslant m \leqslant \infty$), and that the alternative distributions of the sample point admit density functions with respect to a fixed σ-finite measure. It is shown by an example that the restriction to bounded estimates is essential to conclusion (i).

2. INTRODUCTION

Let X be a sample space of points x, and suppose that x is distributed in X according to some unknown one of a given set P of probability measures p. Let g be a real parameter, that is, a real valued function on P. g is said to be estimable if there exists at least one unbiased estimate of g, say $t(x)$, such that the variance of t is finite for each p. Suppose that g is estimable. A particular unbiased estimate of g, t_0 say, is said to be efficient at p_0 if the variance of t_0 does not exceed that of any other unbiased estimate when x is distributed according to p_0; t_0 is uniformly efficient if t_0 is efficient at each p in P.

A theory of uniformly efficient estimation was developed by Lehmann and Scheffé (1950, 1955). This theory includes the following application of a theorem of Rao and Blackwell. Suppose that there exists a sufficient statistic, $y = s(x)$ say, such that the set of alternative distributions of y is complete, that is to say, there exist no unbiased estimates of zero that depend on x only through s except for the trivial estimate $t(y) \equiv 0$. Such a statistic is called a complete sufficient statistic. In this case, according to Theorem 5.1 of Lehmann and Scheffé (1950), every estimable g admits a uniformly efficient estimate; moreover, the uniformly efficient estimate of a given g can be characterised as the unbiased estimate that depends on x only through s and is of finite variance for each p, or, alternatively, as the conditional expectation of t given y, where t is any unbiased estimate such that the variance of t is finite for each p.

As may be seen from examples (cf., Girshick, Mosteller and Savage (1946); Halmos (1946); Rao (1946, 1949); Lehmann and Stein (1950); Lehmann and Scheffé (1950, 1955)) the theorem just stated provides a powerful technique for the discovery (or verification, in case a likely estimate is on hand) of the uniformly efficient estimate of a given g. It can also be seen from examples, however, that this technique is not always available, that is to say, a complete sufficient statistic may not exist in a given case (cf., Lehmann and Scheffé (1950); Lehmann and Stein (1950)). The main object of this paper is to show, under certain technical qualifications, that this technique *is* available whenever *every* estimable g admits a uniformly efficient estimate. Some related conclusions, which are trivially true if a complete sufficient statistic exists, but which happen to be valid in general, are also established (cf., para 1 of Section 1).

One of the qualifications referred to above is necessitated by the possible existence of unbiased estimates of zero that are of infinite variance. This difficulty is excluded simply by letting completeness of a statistic y mean that there exist no nontrivial unbiased estimates of zero that depend on x only through y and are of finite variance for each p. It is easily seen that this definition of completeness is adequate for application to minimum variance estimation theory. Another source of difficulty is that it may be impossible to carry out the maximum reduction of the given sample space by means of sufficient statistics (cf., Pitcher, 1957). This difficulty is removed by assuming that (*) X is, or may be taken to be, a Borel set of the m-dimensional Cartesian space ($1 \leqslant m \leqslant \infty$), and that each p in P admits a probability density function with respect to a fixed σ-finite measure. This assumption is valid in most experiments of statistical interest, including sequential ones.

The following Section 3 discusses efficiency at a point. A geometric characterisation of estimates efficient at a given point is obtained (Theorem 1). It is pointed out that this characterisation yields necessary and sufficient conditions (a) in order that the Rao-Blackwell method of improving a given estimate always lead to the efficient estimate, and (b) in order that the sequence of Bhattacharyya lower bounds to the variance always tend to the exact bound. These conclusions concerning efficiency at a point are used in Sections 4 and 5 to establish the main theorems of the paper (Theorems 4 and 5).

The writer is indebted to his colleague C. R. Rao for pointing out that certain propositions closely related to the ones established here are treated in (Rao, 1952). An argument used in the paper just cited has enabled the writer to simplify the proofs of the main theorems of this paper.

In the formal exposition of the following sections it is convenient and of some theoretical advantage to discuss subfields rather than statistics or measurable transformations. Such technical terms as are used without explanation are defined in the first part of (Bahadur, 1954.) The relation between transformations, statistics, and subfields is discussed, for example, in (Bahadur and Lehmann, 1955) and in (Bahadur, 1955a), and conclusions concerning subfields established here can be formulated in other terms (*e.g.* as in Section 1) by reference to these papers. In particular, if (*) holds, then to all intents and purposes a subfield corresponds to a statistic, and an estimate measurable with respect to a subfield is an estimate that depends on the sample point only through the corresponding statistic.

3. Efficiency at a point

Let X be a set of points x, S a field of subsets of X, and P a set of probability measures p on S. The framework X, S, P will remain fixed throughout this section and the following ones. A *parameter* is a real valued function on P. An *estimate* is defined (without reference to any particular parameter) to be a real valued S-measurable function of x. An estimate t is an *unbiased* estimate of a parameter g if t is P-integrable and

$$E_p(t) = g(p) \text{ for each } p \text{ in } P, \qquad \ldots \quad (3.1)$$

where E_p denotes expected value when p obtains.

In order to simplify the writing, this section and the following ones are written as if the empty set were the only S-measurable set of p measure zero for each p in P. If this condition is not satisfied in the given case, many of the definitions, arguments and conclusions to follow should be accompanied, strictly speaking, by a null set statement or qualification. For example, the assertion that t_1 is the 'only' estimate with a certain property is likely to mean that if t_2 also has the property then $p(t_2 = t_1) = 1$ for each p in P. Specifically, the following convention is observed throughout the paper. If A and B are sets in S, $A = B$ means that $(A-B)\bigcup(B-A)$ is p-null for each p in P; if t_1 and t_2 are estimates, $t_1 = t_2$ means that $\{x : t_1(x) = t_2(x)\}$ = the whole space X. The relations of inclusion and equality between classes of sets of X, and between classes of estimates, are to be interpreted, of course, in terms of this convention. The interested reader may verify that in this paper the adoption of this convention does not lead to any real difficulty, partly because much of the discussion is in terms of subfields.

It is assumed in this section that P is a dominated set. Let there be given a measure μ on S such that the following three conditions are satisfied: μ is σ-finite; each p in P is absolutely continuous with respect to μ; and each of the density functions $dp/d\mu$ is square integrable with respect to μ, that is, $\int_X (dp/d\mu)^2 d\mu < \infty$ for each p in P. In most applications, μ is a probability measure in P, or perhaps a mixture of the measures in P, but the following development is valid provided only that μ and P satisfy the conditions stated.

Let V denote the real linear space of all estimates t such that $\int_X t^2 d\mu < \infty$. For $t_1, t_2 \in V$ write $(t_1, t_2) = \int_X t_1 \,.\, t_2 d\mu$ and $\| t_1 \| = (t_1, t_1)^{\frac{1}{2}}$. We have

$$E_p(t) = (t, dp/d\mu) \text{ for all } t \in V, p \in P. \quad \ldots \quad (3.2)$$

Let W denote the subspace ($\equiv$ closed linear manifold) spanned by the set $\{dp/d\mu : p \in P\}$.

Let us say that a parameter g is *μ-estimable* if there exists a $t \in V$ that satisfies (3.1). According to (3.2), g is μ-estimable if and only if there exists a $t \in V$ such that $(t, dp/d\mu) \equiv g(p)$. For any μ -estimable g let U_g denote the class of all estimates $t \in V$ that are unbiased estimates of g. It is easily seen that U_g is a closed and convex subset of V.

An estimate t_0 is said to be a *μ-efficient* estimate of g if $t_0 \in U_g$ and $\|t_0\| = \inf \{\|t\| : t \in U_g\}$. Note that in the special case when μ is a probability measure in P, a μ-efficient estimate is simply an unbiased estimate of minimum variance at μ. Also, in this case, the set W is the subspace generated by the likelihood ratios relative to μ.

Theorem 1 below asserts the existence and uniqueness of the μ-efficient estimate, and gives a geometrical description of it. The mathematical content of the theorem is virtually the same as that of the principal theorem of (Stein, 1950), (cf. also Section 8 of Barankin, 1949), but the present statement and proof seem simpler. A similar treatment was given earlier by Basu (1953).

Theorem 1: *Let g be a μ-estimable parameter.* (i) *The set $U_g \cap W$ consists of one estimate, t_0 say.* (ii) *$t_0 = \Pi t$ for every $t \in U_g$, where Π is the orthogonal projection to W.* (iii) *t_0 is μ-efficient.* (iv) *t_0 is the only μ-efficient estimate of g, that is, if $t \in U_g$ and $t \neq t_0$ then $\|t_0\| < \|t\|$.*

Proof: Since Π is a projection, we have

$$(\Pi t, u) = (t, \Pi u) \qquad \dots \quad (3.3)$$

and

$$\| \Pi t \| \leqslant \| t \|, \qquad \dots \quad (3.4)$$

with equality in (3.4) only if $\Pi t = t$. Moreover, since Π is the projection to W,

$$W = \{t : t \in V, \Pi t = t\} = \{\Pi t : t \in V\} \qquad \dots \quad (3.5)$$

(cf., Halmos (1951)).

Choose and fix a $t \in U_g$ and define

$$t_0 = \Pi t. \qquad \dots \quad (3.6)$$

Then $t_0 \in W$, by (3.5) and (3.6). Also , for each p,

$$\begin{aligned} (t_0, dp/d\mu) &= (\Pi t, dp/d\mu) \text{ by } (3.6) \\ &= (t, \Pi\, dp/d\mu) \text{ by } (3.3) \\ &= (t, dp/d\mu) \text{ by } (3.5) \\ &= g(p) \text{ since } t \in U_g, \qquad \dots \quad (3.7) \end{aligned}$$

so that $t_0 \in U_g$. Thus $t \in U_g$ and (3.6) imply $t_0 \in U_g \cap W$, and $\| t_0 \| \leqslant \| t \|$, with equality only if $t_0 = t$ (cf., (3.4), (3.6)).

Since U_g is non-empty by hypothesis, this last conclusion implies, in particular, that $U_g \cap W$ is non-empty. We now show that $U_g \cap W$ cannot contain more than one estimate. Let t_0 and t_1 be functions in $U_g \cap W$. Then $t_1 - t_0 = u$ say, is a function in W such that $(u, dp/d\mu) = (t_1, dp/d\mu) - (t_0, dp/d\mu) = g(p) - g(p) = 0$ for each p in P, that is, u is orthogonal to each $dp/d\mu$; hence u is orthogonal to each t in W; in particular, u is orthogonal to itself, so that $u = 0$. This establishes part (i) of Theorem 1. Parts (ii), (iii) and (iv) are immediate consequences of part (i) and the conclusion of the preceding paragraph. This completes the proof of Theorem 1.

Remark 1 : Suppose that Π is the projection, not to W as in the statement and proof of Theorem 1, but to a subspace containing W. It can then be seen from the above proof that the following is still true : if $t \in U_g$, then $t_0 = \Pi t \in U_g$ and $\| t_0 \| \leqslant \| t \|$, with strict inequality unless $t = t_0$.

A method of improving a given unbiased estimate, due to Rao (1945) and Blackwell (1947), is to replace it by its conditional expectation given a suitable statistic. This method is a special case of the one mentioned in remark 1 above. To be precise, suppose $\mu(X) < \infty$, and let S_0 be a subfield of S such that each $dp/d\mu$ (and therefore each t in W) is an S_0-measurable function. For any S-μ-integrable function t let Ct denote the conditional expectation of t given S_0, that is, Ct is the unique S_0-μ-integrable function such that $\int_A t d\mu = \int_A Ct d\mu$ for every S_0-measurable A. It follows easily from well known properties of conditional expectation that, regarded as an operator on V, C is the orthogonal projection to the subspace of S_0-measurable estimates (cf. Moy (1954), Bahadur (1955b)). Since this last subspace contains W, remark 1 applies to C.

It is interesting to examine whether the Rao-Blackwell method not only improves a given estimate but actually yields the μ-efficient one. The preceding discussion shows that this last will be the case, for every μ-estimable g, if and only if W is the class of all S_0-measurable functions in V. Consequently, in order that the Rao-Blackwell method always yields the μ-efficient estimate, S_0 must be the smallest field such that each $dp/d\mu$ is S_0-measurable. This necessary condition is, however, insufficient. Some necessary and sufficient conditions are given in (Bahadur, 1955b). One of these conditions may be stated as follows. Let us say that W is *algebraic* if it contains every constant estimate, and if for any bounded estimate t in W, t^2 is also in W; and that W is *bounded* if the set of bounded estimates in W is everywhere dense (in the L_2 sense) in W. Then, with S_0 the smallest field such that each $dp/d\mu$ is S_0-measurable, W is the set of all S_0-measurable estimates in V if and only if W is algebraic and bounded.

Remark 2 : As a counterpart to remark 1 above, let Π be the projection to a subspace of W. It then follows from Theorem 1 that corresponding to each μ-estimable g there exists a function in the subspace, t^* say, such that $\Pi t = t^*$ for each $t \in U_g$. Consequently, $\|t\| \geqslant \|t^*\|$ for each $t \in U_g$, the inequality being strict unless $t = t^*$.

The preceding remark describes a general method of obtaining lower bounds to the variance of unbiased estimates, of which various bounds in the literature are special cases. In particular, in case $P = \{p_\theta : \theta \text{ in a real interval}\}$, $dp_\theta = f(x, \theta) d\lambda$ where λ is a σ-finite measure, and $\mu = p_{\theta_0}$, then under suitable regularity conditions (cf., Stein (1950)),the n-th Bhattacharyya bound to the variance, $b_n(g)$ say, may be obtained from remark 2 by letting Π be the projection to the subspace spanned by $h_0, h_1, \ldots,$ and h_n, where $h_r(x) = r$-th partial derivative of $f(x, \theta)/f(x, \theta_0)$, evaluated at $\theta = \theta_0$, $r = 0, 1, \ldots, \ldots n$. It follows from the derivation outlined here that $\lim_{n\to\infty} b_n(g) =$ variance of the θ_0-efficient estimate of g(when θ_0 obtains), for every θ_o-estimable g, if and only if each likelihood ratio $f(x, \theta)/f(x, \theta_0)$ is in the subspace spanned by $\{h_r : r = 0, 1, 2, \ldots\}$. The reader may verify that this condition is always satisfied if we can write $f(x, \theta) \equiv \alpha(x) \, . \, \beta(\theta). \exp[\theta s(x)]$, $-\infty < \theta < \infty$.

4. UNIFORM EFFICIENCY

In this section we consider the framework $X = \{x\}$, S, and $P = \{p\}$ of the preceding section. It is not assumed for the present, however, that P is a dominated set.

Let U denote the class of all S-measurable estimates $t(x)$ such that

$$E_p(t^2) < \infty \text{ for each } p \in P. \quad \ldots \quad (4.1)$$

A parameter g is said to be *estimable* if U contains an unbiased estimate of g; g is *boundedly estimable* if there exists a bounded unbiased estimate of g. Clearly, each boundedly estimable parameter is estimable. If g is estimable, a particular unbiased estimate of g, t_0 say, is *uniformly efficient* if for any unbiased $t \epsilon U$ we have

$$E_p(t_0^2) \leqslant E_p(t^2) \text{ for each } p \text{ in } P.$$

Now let S_0 be an arbitrary subfield of S. We shall say that S_0 is *complete* if $t = 0$ is the only S_0-measurable unbiased estimate of zero in U; S_0 is boundedly complete if $t = 0$ is the only bounded S_0-measurable unbiased estimate of zero.

Theorem 2: *If there exists a sufficient and complete subfield, then every estimable parameter admits a uniformly efficient estimate.*

The proof of this theorem is virtually the same as that of Theorem 5.1 of Lehmann and Scheffé (1950) and so is omitted. It follows from the omitted proof that if S_0 is sufficient and complete, the uniformly efficient estimate of a given estimable g is unique, and can be described as the S_0-measurable unbiased estimate in U, or alternatively, as Ct, where t is any estimate satisfying (3.1) and (4.1) and C denotes conditional expectation given S_0.

The above descriptions of the uniformly efficient estimate suggest that there can be at most one subfield S_0' that is sufficient and complete. It can be seen from the following theorem that this is indeed the case, since the necessary and sufficient subfield, if it exists, is unique (cf. Bahadur, 1954).

Theorem 3: *If S_0 is sufficient and boundedly complete, then S_0 is necessary.*

It may be noted here that this theorem does not presuppose the existence of the necessary and sufficient subfield.

Proof: Let S^0 be an arbitrary but fixed sufficient subfield. We have to show that $S_0 \subset S^0$. Consider a set $A \epsilon S_0$. Let χ denote the characteristic function of A, that is, $\chi = 1$ on A and $= 0$ on $X-A$. Let ϕ be the conditional expectation of χ given S^0, and ψ the conditional expectation of ϕ given S_0. Then

$$0 \leqslant \phi \leqslant 1, \quad 0 \leqslant \psi \leqslant 1 \qquad \dots (4.2)$$

and

$$E_p(\psi) = E_p(\phi) = E_p(\chi) = p(A) \text{ for } p \epsilon P. \qquad \dots (4.3)$$

Write $t = \chi - \psi$. Then t is a bounded unbiased estimate of zero, by (4.2) and (4.3), and t is S_0-measurable since χ and ψ are. Consequently, by bounded completeness of S_0, $t = 0$, that is, $\chi = \psi$. Since $\chi^2 = \chi$, this last relation implies

$$\chi \cdot \psi = \chi. \qquad \dots (4.4)$$

It now follows from (4.4), the definition of ψ, and the S_0-measurability of χ, by a well-known property of conditional expectation, that

$$E_p(\chi) = E_p(\chi \cdot \psi) = E_p(\chi \cdot \phi) \text{ for } p \in P. \qquad \ldots \quad (4.5)$$

We see from (4.3) and (4.5) that

$$E_p[\chi(1-\phi)] = E_p[(1-\chi)\phi] = 0 \text{ for } p \in P. \qquad \ldots \quad (4.6)$$

It follows from (4.2) and (4.6) that $\phi = \chi$. Since ϕ is S^0-measurable by the definition of ϕ, χ is S^0-measurable, so that A is in S^0. Thus $A \in S_0$ implies $A \in S^0$. This completes the proof of Theorem 3.

The following theorem is a converse of Theorem 3.

Theorem 4: *Suppose that P is a dominated set. If every boundedly estimable parameter admits a uniformly efficient estimate, then a sufficient and complete subfield exists.*

The proof of this theorem (as also of Theorem 5 below) is postponed to the following section.

In the general case, let T denote the class of all uniformly efficient estimates, i.e. $t \in T$ if and only if $t \in U$ and t is the uniformly efficient estimate of g defined by (3.1). Now, if a complete sufficient subfield exists, T can be characterised as the class of all estimates in U that are measurable with respect to this subfield. This suggests that in general (irrespective of whether a complete sufficient subfield exists) T can perhaps be characterised as the class of all estimates in U that are measurable with respect to some (not necessarily sufficient) subfield. A related conjecture is that if t is in T, and $u \in U$ is a function of t, then u is in T. It is shown in Section 6 by an example that neither conjecture is valid in general. The following theorems show, however, that both conjectures are 'nearly' valid. The near validity of the second conjecture was demonstrated earlier by Rao (1952).

Let T_b denote the class of bounded estimates in T. Note that T_b is non-empty, since it certainly contains every constant.

Theorem 5: *There exists a subfield S_0 such that (i) T_b is the class of all bounded S_0-measurable estimates, (ii) every S_0-measurable estimate in U is also in T, and (iii) S_0 is necessary and complete.*

Theorem 6: *If $\phi(z_1, z_2, \ldots z_k)$ is a Borel measurable function of k real variables, and if $u = \phi(t_1, t_2, \ldots, t_k)$ is in U, where $t_1, t_2, \ldots$ and t_k are in T_b, then u is in T.*

Proof: Since the t_i are S_0-measurable by Theorem 5, and ϕ is Borel measurable, it follows that u is an S_0-measurable function in U, and so $u \in T$ by Theorem 5. This completes the proof. It is clear that Theorem 6 and its proof can be generalised, e.g. to pointwise limits of functions of estimates in T_b.

In connection with Theorems 5 and 6 it may be worth while to note that if X is a finite set then $T_b = T$. A less trivial (and perhaps less useful) sufficient condition for the same relation is that P be a finite set.

5. PROOF OF THEOREMS 4 AND 5

We shall first establish Theorem 5. It will be seen that a continuation of the same argument then yields Theorem 4. It is convenient to present the argument as a series of propositions, as follows.

The classes U, T, and T_b of estimates are defined in the preceding section. Let S_0 be the smallest field of sets of X such that each estimate in T_b is S_0-measurable. Further, let N_1 denote the class of all unbiased estimates of zero, and N_2 the subclass of all unbiased estimates of zero that are also in U.

Proposition 1 : An estimate $t \epsilon U$ is in T if and only if $t.z \epsilon N_1$ for every $z \epsilon N_2$.

Proof : For a proof of this well known and important proposition see, for example, Lehmann and Scheffé (1950), Theorem 5.3.

Proposition 2 : T_b is an algebraic linear manifold.

Proof : We have to verify that (i) T_b contains the estimate $t(x) \equiv 1$, (ii) $u, v \epsilon T_b$ implies $\alpha u + \beta v \epsilon T_b$ for all constants α and β, and (iii) $t \epsilon T_b$ implies $t^2 \epsilon T_b$. These verifications can be based on proposition 1. For example, to establish (iii) consider a $z \epsilon N_2$. Then $t \, . \, z$ is in N_1, by proposition 1. Since t is bounded, it follows that $t \, . \, z$ is in fact in N_2. Hence $t \, . \, (t \, . \, z) = t^2 \, . \, z$ is in N_1, by proposition 1. Since $z \epsilon N_2$ is arbitrary, it follows from proposition 1 that $t^2 \epsilon T$, and so $t^2 \epsilon T_b$, since t^2 is bounded. This proof of (iii) is due to Rao (1952).

Now choose and fix a finite measure μ on S, and for any S-measurable function $f(x)$ let $\| f \| = (\int_X f^2 \, d\mu)^{\frac{1}{2}}$.

Proposition 3 : If f is S_0-measurable, and $\| f \| < \infty$, there exists a sequence $\{t_n\}$ in T_b such that $\lim_{n \to \infty} \|f - t_n\| = 0$.

Proof : For the purposes of this proof only, let V denote the normed linear space of all S_0-measurable f with $\|f\| < \infty$, with the usual identification of functions that differ on a set of μ-measure zero. Since μ is an arbitrary measure, this last identification is not necessarily the same as the one used elsewhere in this paper. At any rate, since μ is a finite measure, and since each t in T_b is a (literally) bounded S_0-measurable function, we have $T_b \subset V$. Let Y be the closure of T_b. We have to show that $Y = V$.

Since T_b is a linear manifold of bounded functions, including the constants (cf. proposition 2), it is clear that Y is a bounded subspace, and that Y contains the constants. We shall now show that if f is a bounded function in Y then f^2 is in Y. Given a bounded f in Y, let $\{t_n\}$ be a sequence in T_b such that

$$\lim_{n \to \infty} \| t_n - f \| = 0. \qquad \ldots \quad (5.1)$$

Let k be a positive integer, and let c_k be a constant such that $|t_k(x)| \leqslant c_k$. We then have $\|t_k t_n - t_k f\| \leqslant c_k \|t_n - f\|$ for each n. Hence, by (5.1),

$$\lim_{n \to \infty} \| t_k t_n - t_k f \| = 0. \qquad \ldots \quad (5.2)$$

Since T_b is an algebraic linear manifold (cf. proposition 2), and since $t_k t_n = \{(t_k+t_n)^2 - t_k^2 - t_n^2\}/2$, we have $t_k t_n \in T_b$. It now follows from (5.2) that

$$t_k f \in Y \qquad \ldots \quad (5.3)$$

for every k. Let c be a constant such that $|f(x)| \leqslant c$. We then have $\|t_k f - f^2\| \leqslant c\|t_k - f\|$ for every k. Hence

$$\lim_{k \to \infty} \|t_k f - f^2\| = 0 \qquad \ldots \quad (5.4)$$

by (5.1). It follows from (5.3) and (5.4), as desired, that $f^2 \in Y$.

Thus Y is an algebraic and bounded subspace. It follows (Bahadur, 1955b) that there exists a subfield of S_0, say S^0, such that Y is μ-equivalent to the set of all S^0-measurable functions f with $\|f\| < \infty$. Since $T_b \subset Y$, it follows from the definition of S_0 that this S^0 must be μ-equivalent to S_0, and hence $Y = V$. This completes the proof of proposition 3.

Proposition 4 : If $t \in U$, and t is S_0-measurable, then $t \in T$.

Proof: Let t be an S_0-measurable function in U. Choose and fix a p in P and a z in N_2. Since $E_p(t^2) < \infty$, it follows from proposition 3 with $\mu = p$ that there exists a sequence $\{t_n\}$ in T_b such that $E_p(t_n - t)^2 \to 0$. Since $E_p(z^2) < \infty$, it follows that $E_p(t_n z) \to E_p(tz)$. However, $E_p(t_n z) = 0$ for each n, by proposition 1. Hence $E_p(tz) = 0$. Since $p \in P$ and $z \in N_2$ are arbitrary, it follows from proposition 1 that $t \in T$, and this completes the proof.

Proposition 5 : S_0 is necessary and complete.

Proof: Let A be a set in S_0 and let χ denote the characteristic function of A. Then $\chi \in T$ by proposition 4. Consequently, if S^0 is any sufficient subfield, χ must be S^0-measurable, for otherwise the conditional expectation of χ given S^0 would be a different and better unbiased estimate of $g(p) \equiv p(A)$, by the theorem of Rao and Blackwell. Hence $A \in S^0$ whenever S^0 is sufficient. Since $A \in S_0$ is arbitrary, it follows that S_0 is necessary. To show that S_0 is complete, let $t \in U$ an S_0-measurable unbiased estimate of zero. Then t is the uniformly efficient estimate of zero, by proposition 4. Hence $t = 0$. This completes the proof.

We may note here that Theorem 5 is an immediate consequence of the definition of S_0 and propositions 4 and 5. It remains therefore to establish Theorem 4. This will be done by showing that if P is dominated, and T contains an unbiased estimate of every boundedly estimable parameter, then S_0 is sufficient.

Let Q be a countable subset of P such that Q is equivalent to P, i.e. $p(A) = 0$ for all p in Q implies $p(A) = 0$ for all p in P. The existence of such a set Q is assured by Lemma 7 of Halmos and Savage (1949). Let $Q = \{p_1, p_2, \ldots\}$ be an enumeration of Q, let $c_1, c_2, \ldots$ be constants such that

$$c_i > 0, \ \sum_i c_i < \infty \qquad \ldots \quad (5.5)$$

and define

$$\mu(A) = \sum_i c_i p_i(A) \qquad \ldots \quad (5.6)$$

for A in S. We shall then say that μ is a pivotal measure based on Q.

Proposition 6 : For each q in Q, $dq/d\mu$ is S_0-measurable.

Proof: Choose and fix a q in Q. It follows from (5.5) and (5.6) that we can write $\mu = cq+\lambda$ where c is a positive constant and λ is a measure. Hence $1 = c\,(dq/d\mu) + (d\lambda/d\mu)$, and so $0 \leqslant dq/d\mu \leqslant (1/c)$. Thus $dq/d\mu$ is a bounded measurable function.

Write $f = dq/d\mu$, and define $g(p) = E_p(f)$ for p in P. Then g is a boundedly estimable parameter. Consequently, by hypothesis, there exists an unbiased estimate of g in T, say t. In particular, we have $E_p(t) = E_p(f)$, $E_p(t^2) \leqslant E_p(f^2)$ for all p in P. These relations imply, using (5.5) and (5.6), that

$$E_p(t) = E_p(f) \text{ for } p \in Q; \quad \int_X t^2 d\mu \leqslant \int_X f^2 d\mu. \qquad \ldots \quad (5.6)$$

Suppose for the moment that p is restricted to the set Q. Since $dp/d\mu$ is bounded for each p in Q by the first paragraph of this proof, and since μ is a finite measure, Theorem 1 of Section 3 can be applied to determine μ-efficient estimates of μ- estimable parameters on Q. This application shows that $f = dq/d\mu$ is μ-efficient. Consequently, by (5.6), t is also the μ-efficient estimate of the same parameter. Hence $t = f$, by uniqueness of the μ-efficient estimate. Thus f is a bounded estimate in T, i.e. $f \in T_b$. It now follows from the definition of S_0 that $f = dq/d\mu$ is S_0-measurable, and the proof is complete.

Proposition 7 : S_0 is sufficient.

Proof: Let p_1 and p_2 be two measures in P. Choose and fix a countable equivalent subset Q that includes p_1 and p_2 and let μ be a pivotal measure based on Q. It follows from proposition 6 that $dp_i = f_i(x)\, d\mu$, where f_i is S_0-measurable, $i = 1, 2$.

Hence, by the factorization theorem for sufficient subfields, S_0 is sufficient for the set $\{p_1, p_2\}$.

Since p_1 and p_2 are arbitrary measures in P, it follows that S_0 is pairwise sufficient for P. Proposition 7 now follows from the theorem of Halmos and Savage (1949) that in the dominated case pairwise sufficiency is equivalent to sufficiency.

This completes the proof of Theorem 4. The proof shows that the conclusion of the theorem holds provided only that every parameter of the form $g(p) = E_p(dq/d\mu)$ admits a uniformly efficient estimate, where μ is a pivotal measure based on Q, q is a measure in Q, and Q is a countable equivalent subset of P.

6. AN EXAMPLE

It is the object of this section to show by an example that a real valued function of a uniformly efficient estimate is not necessarily uniformly efficient. It follows from the theory developed in Sections 4 and 5 that in any such example a complete sufficient statistic does not exist; the class T of all uniformly efficient estimates cannot be characterised as the class of all square integrable functions of some statistic; there exist estimates in T that cannot be approximated (in any reasonable sense) by bounded estimates in T; and both the sample space X and the set P of alternative distributions on X must be infinite.

To construct such an example, let X be the set of points $x = 0, 1, 2, \ldots$ ad inf., and S the field of all sets of X. We shall choose the set P of alternative distributions of x so that $T =$ the class of all linear functions of x. Such a choice of P evidently furnishes the required example.

Let m be a probability distribution such that

$$E_m(x^2) < \infty, E_m(x^4) = \infty \quad \ldots \quad (6.1)$$

and such that

$$E_m(z_1) = E_m(z_2) = 0 \quad \ldots \quad (6.2)$$

where

$$z_1(x) = (-1)^x . x^2 \text{ for } x = 0, 1, 2, \ldots \quad \ldots \quad (6.3)$$

and

$$z_2(x) = \begin{cases} 1 & \text{if } x = 0 \\ (-1)^x . x & \text{if } x = 1, 2, \ldots \end{cases} \quad \ldots \quad (6.4)$$

It is not difficult to see that such a distribution m exists. Let $p_0 = m$. Next, let

$$p_1(x) = \begin{cases} 2/7 & \text{if } x = 0 \\ 4/7 & \text{if } x = 1 \\ 1/7 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases} \quad \ldots \quad (6.5)$$

and for $k = 2, 3, \ldots$ let

$$p_k(x) = \begin{cases} k(k+1)/4k^2-2 & \text{if } x = k-1 \\ 2(k^2-1)/4k^2-2 & \text{if } x = k \\ k(k-1)/4k^2-2 & \text{if } x = k+1 \\ 0 & \text{otherwise.} \end{cases} \qquad \ldots \quad (6.6)$$

Having thus defined p_θ for integral values of θ in $[0, \infty)$, extend the definition by linear interpolation. Specifically, for any θ with $0 \leqslant \theta < \infty$, let $k = k(\theta)$ be the greatest integer $\leqslant \theta$, let $\alpha = \alpha(\theta)$ be determined by $\alpha k+(1-\alpha)(k+1) = \theta$, $0 < \alpha \leqslant 1$, and define $p_\theta(x) = \alpha p_k(x)+(1-\alpha)p_{k+1}(x)$. Let $P = \{p_\theta : 0 \leqslant \theta < \infty\}$.

To determine the class of uniformly efficient estimates in the framework X, S, P defined above, let the classes N_1 and N_2 of unbiased estimates of zero be defined as in the second paragraph of Section 5. It follows easily from (6.3), (6.4), (6.5) and (6.6) that $E_\theta(z_i) = 0$ for $\theta = 0, 1, 2, \ldots$ and $i = 1, 2$. Hence $E_\theta(z_i) = 0$ for all θ, $i = 1, 2$. Hence az_1+bz_2 is in N_1, for all constants a and b. Now consider an arbitrary z_3 in N_1. Let a and b be constants (possibly both zero) such that with $z_4 = z_3-az_1-bz_2$ we have $z_4(x) = 0$ for $x = 0$ and $x = 1$; it is easily seen from (6.3) and (6.4) that such constants exist. Then z_4 is an unbiased estimate of zero, and z vanishes for $x = 0$ and $x = 1$. It now follows from (6.5) and (6.6) that we must have $z_4(x) = 0$ for all x. Hence $z_3 = az_1+bz_2$.

We conclude that N_1 is the class of all estimates of the form az_1+bz_2. It follows hence by (6.1), (6.3) and (6.4) that N_2 is the class of all estimates of the form bz_2. Since z_2 is never zero, it now follows from proposition 1 that T is the class of all estimates of the form az_1+bz_2/z_2, i.e. of the form $\alpha+\beta x$, since $z_1/z_2 \equiv x$. This completes the verification of the example.

REFERENCES

Bahadur, R. R. (1954): Sufficiency and statistical decision functions. *Ann. Math. Stat.*, **25**, 423–462.

——— (1955a): Statistics and subfields. *Ann. Math. Stat.*, **26**, 490–497.

——— (1955b): Measurable subspaces and subalgebras. *Proc. Am. Math. Soc.*, **6**, 565–570.

Bahadur, R. R. and **Lehmann, E. L.** (1955): Two comments on 'sufficiency and statistical decision functions.' *Ann. Math. Stat.*, **26**, 139–142.

Barankin, E. W. (1949): Locally best unbiased estimates. *Ann. Math. Stat.*, **20**, 477–501.

Basu, D. (1953): *Some Contributions to The Theory of Statistical Inference.* D. Phil. thesis, University of Calcutta (unpublished).

Blackwell, D. (1947): Conditional expectation and unbiased sequential estimation. *Ann. Math. Stat.*, **18**, 105–110.

Girshick, M. A., **Mosteller, F.** and **Savage, L. J.** (1946): Unbiased estimates for certain binomial sampling problems with applications. *Ann. Math. Stat.*, **17**, 13–23.

Halmos, P. R. (1946): The theory of unbiased estimation. *Ann. Math. Stat.*, **17**, 34–43.

——— (1951) : *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, New York, Chelsea, 1–48.

HALMOS, P. R. and SAVAGE, L. J. (1949) : Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.*, **20**, 225–241.

LEHMANN, E. L. and STEIN, C. (1950) : Completeness in the sequential case. *Ann. Math. Stat.*, **21**, 376–385.

LEHMANN, E. L. and SCHEFFÉ, H. (1950) : Completeness, similar regions, and unbiased estimation. *Sankhyā*, **10**, 305–340.

——— (1955) : Completeness, similar regions, and unbiased estimation. *Sankhyā*, **15**, 219–236.

SHU-TEH CHEN MOY, (1954) : Characterisations of conditional expectation as a transformation on function spaces. *Pacific Journal of Mathematics*, **4**, 47.

PITCHER, T. S. (1957) : Statistical problems that do not admit necessary and sufficient statistics or subfields. *Ann. Math. Stat.*, **28**, 267–268.

RAO, C. R. (1945) : Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, **37**, 81–91.

——— (1946) : Minimum variance and the estimation of several parameters. *Proc. Camb. Phil. Soc.*, **43**, 280-283.

——— (1949) : Sufficient statistics and minimum variance estimates. *Proc. Camb. Phil. Soc.*, **45**, 213–218.

——— (1952) : Some theorems on minimum variance estimation. *Sankhyā*, **12**, 27–42.

STEIN, C. (1950) : Unbiased estimates of minimum variance. *Ann. Math. Stat.*, **21**, 406–415.

Paper received: September, 1957.

THE NATIONAL SAMPLE SURVEY

NUMBER 9

REPORT ON THE SAMPLE SURVEY OF DISPLACED PERSONS IN THE URBAN AREAS OF THE BOMBAY STATE

JULY—SEPTEMBER 1953

BY
SATYABRATA SEN

WITH A FOREWORD BY
P. C. MAHALANOBIS

THE NATIONAL SAMPLE SURVEY

NUMBER 9

REPORT ON THE SAMPLE SURVEY OF DISPLACED PERSONS IN THE URBAN AREAS OF THE BOMBAY STATE

JULY—SEPTEMBER 1953

CONTENTS

PAGE

FOREWORD : by P. C. Mahalanobis 229
CHAPTER ONE : Introduction 231
CHAPTER TWO : Summary of results 236
CHAPTER THREE : Demographic statistics 243
CHAPTER FOUR : Expenditure on consumption 254
CHAPTER FIVE : Income 258
CHAPTER SIX : Loans 265
CHAPTER SEVEN : Housing and miscellaneous statistics 272
List of Tables 283
Tables 285
Facsimile of the schedule of investigation ... 327

INDEX TO TABLES IN THE TEXT

PAGE

CHAPTER ONE

TABLE 1.1 : Distribution of sample households by division, district and locality type .. 234

CHAPTER THREE

TABLE 3.1 : Number of sample households and persons per household 243
TABLE 3.2 : Distribution of households by their size 243
TABLE 3.3 : Percentage distribution of households by mother tongue 244
TABLE 3.4 : Percentage distribution of displaced persons of Bombay by states from which they migrated : 1951 population census 245
TABLE 3.5 : Percentage distribution of displaced persons by age groups 245
TABLE 3.6 : Percentage of literacy (urban areas) 246

PAGE
Table 3.7 : Percentage distribution of persons by education standard 247
Table 3.8 : Percentage distribution of earners, earning dependents and non-earning dependents 247
Table 3.9 : Percentage distribution of economically active population by means of livelihood classes 248
Table 3.10 : Percentage distribution of economically active population by employment status .. 249
Table 3.11 : Distribution of earners and earning dependents (non-earning dependents excluded) by livelihood classes with average number of days employed and average income earned per month per person 250
Table 3.12 : Distribution of earners and earning dependents (non-earning dependents excluded) by days of employment in a month with average income per person 251
Table 3.13 : Distribution of households by pre-migration livelihood classes and proportions continuing the same activities during the period of the survey 252

CHAPTER FOUR

Table 4.1 : Comparative consumer expenditure of displaced and general population in rupees per household per month 254
Table 4.2 : Average of working class cost of living index numbers for the two periods, April–September 1952 and July–September 1953 255
Table 4.3 : Percentage distribution of displaced households and general households in Bombay by levels of monthly expenditure 256
Table 4.4 : Patterns of consumption of displaced households in different localities and of general households in urban areas of Bombay 256

CHAPTER FIVE

Table 5.1 : Average monthly receipt and expenditure per household in rupees 258
Table 5.2 : Distribution of monthly receipts per household by sources of receipts 259
Table 5.3 : Number and percentage of households depending partly or wholly on different sources of receipts for meeting their consumption demands 260
Table 5.4 : Number of households receiving and the amount received per household by type of source 261
Table 5.5 : Number of households taking loans for consumption with or without interest by types of securities 262
Table 5.6 : Distribution of households by income to expenditure percentage levels .. 263

CHAPTER SIX

Table 6.1 : Number of households with an amount per household of outstanding loans by types of loans 265
Table 6.2 : Distribution of households by size of outstanding government and non-govenrment loans 266
Table 6.3 : Receipts, utilisation and repayment per household of government loans and grants for different purposes in government colonies 267
Table 6.4 : Receipts, utilisation and repayment per household of government loans and grants for different purposes in other localities 268
Table 6.5 : Receipts, utilisation and repayment per household of government loans and grants for different purposes in all localities 269
Table 6.6 : Distribution of households by the amount of government loans received by them .. 269
Table 6.7 : Distribution of households receiving government loans by percentage of utilisation for purposes for which taken 270
Table 6.8 : Percentage distribution of households receiving trade and professional loans by percentage utilisation for the purpose taken 271

CHAPTER SEVEN

TABLE 7.1 : Distribution of households by types of possession of living accommodation .. 272
TABLE 7.2 : Distribution of households by the number of rooms occupied 272
TABLE 7.3 : Distribution of households by the area of floor space occupied 273
TABLE 7.4 : Distribution of households by type of construction of living accommodation .. 274
TABLE 7.5 : Distribution of households by type of construction of living accommodation .. 275
TABLE 7.6 : Distribution of households by type of lighting 275
TABLE 7.7 : Distribution of households by source of drinking water with average distance of source 276
TABLE 7.8 : Distribution of households by number using the same source of drinking water .. 277
TABLE 7.9 : Distribution of households by source of non-drinking water with average distance of source 277
TABLE 7.10 : Distribution of households by numbers using the same source for non-drinking water 278
TABLE 7.11 : Distribution of households by type of latrines used with average distance of the latrines 278
TABLE 7.12 : Distribution of households by numbers using the same latrines (structures) .. 279
TABLE 7.13 : Number and proportion of households preferring to continue the same means of livelihood as followed at the time of the survey 280
TABLE 7.14 : Distribution of households by states preferred to live in 281
TABLE 7.15 : Distribution of household heads by degree of knowledge of local language .. 282

PRINCIPAL PARTICIPANTS

Design and Planning : SHRI D. B. LAHIRI, *Indian Statistical Institute*
SHRI N. C. GHOSH, ,,

Field Work : PROF. H. GHOSH, *Chief Director*, *NSS* (1953)
SHRI S. KANNAN, *Superintendent, Bombay Block*, *NSS* (1953)
SHRI S. C. DHAR, *Assistant Director, West India Range*, *NSS* (1953)
SHRI H. K. CHATURVEDI, *Indian Statistical Institute*

Processing : SHRI J. M. SEN GUPTA, *Indian Statistical Institute*
SHRI H. B. CHAUDHURY, ,,

THE NATIONAL SAMPLE SURVEY

NUMBER 9

REPORT ON
THE SAMPLE SURVEY OF DISPLACED PERSONS IN THE URBAN AREAS OF THE BOMBAY STATE
JULY—SEPTEMBER 1953

FOREWORD

0.1. The Indian Statistical Institute was requested in January 1953 to carry out a survey of the displaced persons in the Bombay State. It was found from the records of 1951 Census that about 88 per cent of the displaced persons in that State were living in urban areas. It was, therefore, decided to restrict the survey to urban areas of the State.

0.2. The actual survey was carried out from July to September 1953. An advance copy of the draft report, prepared by Shri Satyabrata Sen of the Indian Statistical Institute, was sent to the Ministry of Rehabilitation on 8 February 1954, and the report in its final form was submitted in October 1955.

0.3. The refugee problem arose out of the partition of India in August 1947. Even before the date of partition the movement of population into what later became the Indian Union had started. The influx continued on an increasing scale for sometime after partition, and although the intensity diminished considerably and became intermittent in time, it never ceased altogether. It appears that more people have moved into the Indian territory than out of it.

0.4. Relief had to be provided immediately; but before any long-term plan for rehabilitation could be formulated it was necessary to collect detailed information relating to the immigrants. At the instance of the Government of India a number of surveys were undertaken and carried out in 1947-48 by the Indian Statistical Institute in parts of both eastern and western India. The main purpose of these earlier surveys was to ascertain the number of refugees and to obtain some information—demographic and other—regarding their social and economic conditions. Later on, certain other surveys were also carried out. For example, a "Survey of Economic Conditions of Refugees in West Bengal" was made in January 1953 in a number of refugee colonies to find out the extent of progress of rehabilitation and resettlement. The report was submitted in February 1953 to the Fact-finding Committee (appointed by the Ministry of Rehabilitation, Government of India) of which Shri Satyabrata Sen was a member. The request for the survey covered by the present report came a little later.

0.5. The total number of displaced persons covered by the present survey was 2,94,000. The enquiry was mainly directed to finding out the position of the displaced households in respect of their size, age distribution, extent of literacy, language spoken, income, consumer expenditure, indebtedness, housing conditions, employment status, means of livelihood, preference regarding means of livelihood and place of residence.

0.6. It was found that a little over 50 per cent of households followed the old means of livelihood while the rest had changed their vocation, the largest shift, as could be expected, occurring from agriculture. Preference regarding means of livelihood showed that nearly 62 per cent of the households would prefer to follow the occupation which they had at the time of the survey.

0.7. The present report is not exhaustive. The facsimile schedule will show that a large amount of data were collected on a wide variety of subjects from which a good deal of additional information can be extracted. In preparing this report, a selection was made of topics which were likely to be of immediate need in planning for rehabilitation.

5 August 1957

P. C. Mahalanobis

THE NATIONAL SAMPLE SURVEY

NUMBER 9

REPORT ON
THE SAMPLE SURVEY OF DISPLACED PERSONS IN THE URBAN AREAS OF THE BOMBAY STATE: JULY—SEPTEMBER 1953

This Report on the Sample Survey of Displaced Persons in the Urban Areas of the Bombay State was prepared by the Indian Statistical Institute and is being published in the form in which it was submitted to the Government of India. The views contained in this Report are not necessarily those of the Government of India.

CHAPTER ONE

INTRODUCTION

1.1. The sample survey of displaced persons in the Bombay State was organised by the National Sample Survey (NSS) in July to September 1953 at the request of the Ministry of Rehabilitation, Government of India. As 88 per cent of the displaced population in the State were concentrated in urban areas, the survey was confined to those areas only. The selection of sample households was made on the basis of DP slips of 1951 Population Census. Information regarding social and economic conditions of the displaced households was collected in the survey by interviewing the households.

1.2. The actual collection of data was started at the end of June 1953 with a few investigators who were recruited and trained just for this survey and all of whom were Sindhis. They were later supplemented by trained investigators of the NSS located in the Bombay State when one round of NSS investigation was nearing completion and another was to begin. The design of the survey, the questionnaires and the necessary instructions to the investigators were prepared by the staff of the Indian Statistical Institute where the Technical Wing of the NSS is located. The analysis of the collected data and the drafting of this report were also done in the Institute.

1.3. The information related to various aspects of their lives, namely, (a) demographic and occupational particulars, such as, age and sex of the members of the households, their education standard, economic status, means of livelihood and employment status; (b) earnings of the working members and various kinds of receipts by the members of the households with reference to a period of one month prior to the date of survey; (c) expenditure of the households by various items of consumption in a period of one month preceding the date of survey; (d) loans taken by the households from private and government sources and detailed particulars regarding type of government loans and the extent of their utilisation; (e) housing, sanitation and water supply and also their knowledge about the local language and their preference regarding place of rehabilitation and occupation.

1.4. The sampling was done in two stages. In the first stage, a sample of census tracts was selected and in the second, a sample of displaced households living in these census tracts was investigated. Since all the census tracts where the major government colonies were located were included in the survey, the sample was composed of two types of displaced households, e.g., (a) those living in government colonies and (b) those living in other localities. In all 1349 questionnaires could be completed in the survey and of these 880 households were living in government colonies and 469 in other localities. The results of the survey in this report have also been given with this breakdown of government colonies and other localities.

Sampling design and the work

1.5. The survey was restricted to displaced persons living in the urban areas of the Bombay State. As only 12 per cent of this population were living in rural areas and were scattered, they were excluded from the survey. Their inclusion would have pushed the cost of the survey disproportionately high.

1.6. The Population Census of 1951 took a special count of the displaced population and the census slips were kept apart for special tabulation. These slips were used as the frame for selecting samples. The slips, however, related to individual persons and since the unit of investigation in this survey was a household, only the slips relating to the household heads were required for our purpose. This involved sorting out the slips for the household heads from the rest of the slips for the displaced persons.

1.7. The whole of the Bombay State for the 1951 Census was divided into rural and urban areas and then grouped into about 200 census tracts. For the purpose of this survey, the rural tracts as also the urban tracts having 100 or less displaced persons were excluded. Sample tracts were then selected out of urban tracts having over 100 DPs; and also all the tracts having over 3,000 DPs were included. In all, 24 urban tracts were selected as first stage samples from 43 tracts spread over 15 districts of the State. 294,000 of the total of 298,000 urban displaced persons were living in these 43 tracts and thus, the coverage of the survey was extended to 98.6 per cent of the urban DPs of the State.

1.8. The DP slips of these selected census tracts were then obtained from the Superintendent of Census, Bombay State and the slips for the household-heads were sorted out from the rest. A list was made of the names of the household-heads together with census house numbers and addresses as given in these slips. Sample households were then drawn from these lists with a uniform sampling fraction. There was no way of ascertaining from the slips as to who were living in government colonies and who were living in other localities. But as the samples were selected at random, it was expected that colony and non-colony population would be automatically represented in the sample and in the same relative proportions.

1.9. The preliminary work for the survey was started as early as April 1953. The work involved collection of full details regarding census material from Census Authorities, selection of sample tracts, collection of DP slips from the different Regional Census Offices in the State, sorting of slips to obtain the slips of the household-heads, listing them and finally selecting samples from these lists. This work was completed by the end of June 1953.

1.10. The actual work of data collection was started in Ulhasnagar Township near Kalyan on 30th June 1953. On the first day, the Superintendent of NSS in the Bombay State accompanied by investigators called on the Administrator of the colony and the latter agreed to give all necessary help to the investigators. In addition to this help, the cooperation of the local influential persons and of organisations like the Sindh Seva Samiti was very valuable.

1.11. As the survey proceeded, it was found that locating the sample households on the basis of inadequate addresses as given in the census slips was very difficult. In many localities, the census numbering of the houses was not so well arranged as to be followed easily by the investigators. There were also cases where the houses could be located but the neighbours reported that persons as stated in the slips never lived in them. Many of the houses could not be traced at all. The city of Bombay was the most difficult area. Many of the houses were white-washed and the census numbers were not visible. In some of the slips only the names of the streets and no numbers were given, and hence the investigators had to go from one end of the street to the other with only the name to locate the household. They also made use of all possible sources, such as particulars available in post offices, rationing offices, voters' list etc.

1.12. As a result, although much more time than was expected at the beginning was given to the searching of sample households, the number of households who could not be traced or had gone away from the addresses given in the census slips was very large indeed. Only 1349 out of 2795 or, about 48 per cent of sample households selected from the lists could be investigated. As the displaced persons outside the colonies were living together with the general population, it would be a difficult and costly job to prepare a frame for sample selection containing only the displaced persons. Even knowing that there would be high casualties it was decided that the frame would be based on census slips. When the cost is taken into consideration, there was really no practical way of getting a sampling frame other than the one used in this survey. The amount of work that went in the preliminaries of the survey was higher than was expected. Of a total volume of 86 man-months work, 32 man-months were spent for the preparatory work and 54 man-months for actual investigation.

1.13. Because of this large proportion of casualties, a question arises as to what extent these omissions have vitiated the representativeness of the sample.

In the next chapter, therefore, wherever possible, the results of this survey have been compared with the figures of the 1951 Population Census regarding the displaced persons. Comparatively stable features, such as age-structure, proportion of literacy, occupational distribution etc. do not change appreciably within a period of two or three years and as it will be seen later, the agreement between the two sets of figures is satisfactory.

1.14. A distribution of the sample households by the different divisions of the State, districts, census tracts and locality type is given in Table (1) at the end of the report but a summary of these figures is given in Table (1.1) below.

TABLE (1.1) : DISTRIBUTION OF SAMPLE HOUSEHOLDS BY DIVISION DISTRICT AND LOCALITY TYPE

division	district	number of sample households		
		government colonies	other localities	all localities
(1)	(2)	(3)	(4)	(5)
1. Northern Deccan	E. Khandesh	36	—	36
	Poona	10	50	60
	Nasik	4	27	31
	Ahmednagar	10	16	26
2. Southern Deccan	—	—	—	—
3. Gujrat	Mehsana	1	3	4
	Panch Mahals	1	15	16
	Ahmedabad	138	81	219
	Surat	—	5	[illegible]
	Banaskantha	—	3	[illegible]
	Baroda	13	42	55
4. Greater Bombay	Greater Bombay	53	204	257
5. Konkan	Thana	614	18	632
6. total		880	469	1349

1.15. Of the 1349 sample households, 880 were living in government colonies and 469 outside. The households in the government colonies were concentrated mainly in the districts of Thana, Ahmedabad, Greater Bombay and East Khandesh. The non-colony households were concentrated mainly in Greater Bombay, Ahmedabad, Poona and Baroda.

1.16. The proportion of households living in the government colonies to the total sample households is about 65 per cent. But the proportion of persons living in government colonies to total persons of the sample is 62 per cent (4288 persons of 6914 persons). According to the figures supplied by the Rehabilitation Department of the Bombay Government, the total number living in government colonies in March 1953 was 179,000 persons. This figure is 61 per cent of the total urban displaced population of 294,000 covered by this survey. The agreement between the proportion of the sample and that of the population is very satisfactory.

THE TABLES

1.17. The main Tables with detailed figures are attached at the end of this report. These can be arranged by subjects as follows :

1.	Number and distribution of samples	Table (1)
2.	Demographic statistics	Tables (2) to (10)
3.	Expenditure on consumption	Table (11)
4.	Income	Tables (12) to (16)
5.	Loans	Tables (17) to (24)
6.	Housing and miscellaneous statistics	Tables (25) to (41)

In addition, summary tables have been given in the following five chapters where the more important results of the survey have been discussed in detail.

1.18. A facsimile of the questionnaire used in the survey is attached at the end of the report showing the various items for which the figures have been collected. But the figures on all the items have not been analysed and the results given in this report are by no means exhaustive. Only a selection has been made of those topics in which the Government of India is likely to be specially interested.

Chapter Two

SUMMARY OF RESULTS

Introduction

2.1. The sample survey of displaced persons in the Bombay State was carried out in July to September 1953. 88 per cent of the displaced population having settled in the urban areas of the State, the survey was confined to those areas only. The coverage of the survey extended to 98.6 per cent of the urban displaced persons of the State, *i.e.*, to 294,000 out of a total of 298,000 urban displaced persons.

2.2. The investigation was done by persons specially recruited and trained for this survey all of whom were Sindhis. They were later joined by trained NSS investigators during the interregnum between the completion of one round of NSS investigation and the commencement of another. The design, questionnaires, instructions to investigators, analysis and the report were prepared by the Indian Statistical Institute where the Technical Wing of the NSS is located.

2.3. Information was collected in respect of various aspects of the lives of displaced persons. It related to (a) demographic and occupational particulars, (b) earnings of the working members and various other receipts of the household, (c) expenditure of the households during the month preceding the date of survey, (d) loans of the households from various sources and their utilisation, (e) housing, sanitation, water supply, knowledge of the local language, preference regarding occupation and habitation.

2.4. The sample was composed of two types of displaced households, namely (a) those living in government colonies, and (b) those living in other localities. In all, 1349 questionnaires could be completed of which 880 households were living in government colonies and 469 in other localities. The amount of preliminary work involved was much more than was expected and out of 86 man-months work, 32 man-months were spent for the preparatory work and 54 man-months for actual investigation.

Demographic statistics

2.5. *Household size :* The average size of households of the displaced persons was 5.1 as against nearly 4.7 of the general population according to 1951 Census. The size of households living outside government colonies was of 5.6 persons and of those in government colonies of 4.9 persons.

2.6. *Mother tongue :* 81.0 per cent of the displaced households were Sindhi-speaking, 7.4 per cent were speaking Punjabi, 6.2 per cent spoke Gujrati, 1.4 per cent Marathi and 3.2 per cent other languages. The Sindhi-speaking households in government colonies were 88.3 per cent and in other localities 67.2 per cent. The proportions of Punjabis and Gujratis were greater in non-government colonies.

2.7. *Age :* Children belonging to the age group 0–4 in government colonies and other localities were respectively 12.3 and 11.6 per cent. The distribution in age group 5–14 were about the same in the two areas, it being 27.1 per cent in government colonies and 27.9 per cent in other localities. Persons in the age group 15–54 in the corresponding areas were 52.3 and 53.1 per cent. In the age group 55–74 distribution showed 8.0 per cent in government colonies and 7.0 per cent in other localities.

2.8. *Literacy and education :* The proportion of literacy in government colonies was 51.5 per cent compared with 65.7 per cent in other localities. A higher proportion of males (74.3 per cent) and a still higher proportion of females (56.7 per cent) living outside government colonies were literate. Thus in the case of both males and females, people living in non-government colonies had better educational qualifications. Among the males of all localities 30.9 per cent were illiterate, 42.5 per cent merely literate, 9.3 per cent matriculates and 1.7 per cent were graduates. Among the females of all localities only 2.2 per cent were matriculates of whom 0.4 per cent were graduates. Nearly 90 per cent of females were either illiterate or merely literate.

2.9. *Economic status :* The proportions of earners, earning dependents and non-earning dependents in the colonies and outside were about the same. In government colonies, 18.9 per cent were earners, 3.3 per cent earning dependents and 77.8 per cent non-earning dependents, the corresponding percentages in other localities being 19.5, 3.5 and 77.0.

2.10. *Livelihood classes :* The proportion of active population engaged in agriculture was negligible in both government colonies and other localities, it being 0.3 per cent in the former area and 0.2 per cent in the latter. The proportions in 'other production' were more or less the same for the government colonies (16.8 per cent) and other localities (15.2 per cent) and also the proportions engaged in transport operations were roughly the same in both areas (4.9 per cent in government colonies and 5.3 per cent in other localities). In trade the proportion was somewhat higher among the non-colony population (47.4 per cent) than among those in government colonies (42.1 per cent). The proportion of employed persons in other services was higher among population in government colonies (35.9 per cent) than among those living outside (31.9 per cent). Other and miscellaneous services include personal services, construction work, hotel-keeping, repairing services, professional services and miscellaneous receipts from rent, charity etc.

2.11. *Employment status :* In government colonies, about 98 per cent of the economically active population were either employees or were working on own account. In other localities, about 96 per cent were either employees or working on own account. The proportion of employers, though very small, was significantly higher in other localities (2.8 per cent) than in government colonies (0.7 per cent). A very small proportion lived on unearned incomes.

2.12. *Intensity of employment :* Among all economically active persons of all livelihood classes in government colonies and in other localities a person was employed for a period of 24 days in a month, the average income per person being Rs. 71.8 in the former area and Rs. 118.1 in the latter. The income figures for each livelihood class were also higher in other localities. Over four-fifths of the active population in government colonies and more than nine-tenths in other localities were working for more than 21 days a month. Thus a higher proportion of persons in government colonies was under-employed.

2.13. *Livelihood after migration :* In government colonies, 47.1 per cent of the households continued to follow the same means of livelihood as they did before migration. In other localities, 57.5 per cent pursued their old activities. Taking all localities together, 50.7 per cent of households continued their old vocations. Greatest shift occured in agriculture. Only 0.6 per cent continued to depend on it. Professional services involved the least change in occupation and 67.0 per cent of households remained in the old occupation. 64.2 per cent of households continued to follow trade as before. Nearly 60 per cent of households remained in crafts and cottage industries. Roughly, a half of the households continued to follow the old vocations in construction, transport and services and about a third continued in large industries and other types of odd jobs.

EXPENDITURE ON CONSUMPTION

2.14. *Consumer expenditure :* The per household expenditure of the displaced population of all localities taken together was Rs. 143.2 per month in the period July-September 1953. In government colonies it was Rs. 112.3 compared to Rs. 201.3 in other localities. A comparison of the consumer expenditure of all localities with the per household expenditure of the general population of some urban areas of the Bombay State as revealed by the data of the 4th round of the National Sample Survey (April-September 1952) shows that the two estimates are nearly equal. In spite of this, the position of the displaced households turns out to be somewhat worse when the large size of these households and the increase in cost of living from the time of 1952 survey to that of 1953 survey are taken into consideration.

2.15. *Pattern of consumption :* In government colonies a higher proportion was spent on food, namely 63.7 per cent as compared with 54.8 per cent by households living in other localities. The proportion of expenditure for the residual items was lower in the government colonies (18.4 per cent). The expenditure for house rent in government colonies was also lower, namely 3.0 per cent compared to 9.8 per cent in other localities. The proportion of expenditure on clothing was about the same in both the settlements (8.6 per cent in government colonies and 9.4 per cent in other localities). Taking the households of all localities together, it was found that 59.4 per cent and 9.0 per cent were spent respectively for food and clothing and 19.4 per cent on residual or miscellaneous items.

INCOME

2.16. The monthly receipts were higher than the expenditure in both government colonies and other localities and although the receipts from various sources were meant for consumption they could not be regarded as incomes in the proper sense. In government colonies the monthly money receipt per household was Rs. 123.1 whereas in other localities it was Rs. 225.0.

2.17. Particulars regarding current receipts were collected under such items as (1) income from occupation, (2) receipts from other sources, (3) imputed value of supplies obtained from one's own enterprise, (4) government dole, (5) non-government help and (6) non-government loan. Taking the households of all localities together, the total receipt per household was Rs. 158.5 of which 67.5 per cent were income from occupation and 17.9 per cent from other sources. The proportions for government dole, non-government help and non-government loan were respectively 1.8 per cent, 4.4 per cent and 8.0 per cent. The imputed value of own supply was only 0.4 per cent.

2.18. In government colonies, dependence on government dole was greater than in other localities and 10.2 per cent of households depended on it. In other localities, such households were only 2.3 per cent. The most important source of receipt in both the settlements was the own (residual) income of the households, 51.5 per cent of households in government colonies and 75.0 per cent of households in other localities depending on this source. Taking households of all localities together, 7.5 per cent of households depended on government doles, 9.6 per cent and 18.7 per cent respectively on non-government help and non-government loan, 4.5 per cent on sale of assets and 59.7 per cent on their income alone.

2.19. For all localities taken together the amount of government dole per receiving household was Rs. 39.4. The amount of non-government help per receiving household was Rs. 71.7. Non-government loans were Rs. 67.8 per receiving household for all localities. Similarly the amount obtained by sale of assets was Rs. 76.2 per household. In all cases the per household receipts in other localities were greater than in government colonies.

2.20. In examining how far the incomes of the households could meet their expenditures only the receipts which could be regarded as proper incomes were taken into consideration. Taking the households living in all the localities together, 39.8 per cent of the households had incomes above expenditure, 63.3 per cent had incomes covering more than 90 per cent of their expenditure and 67.5 per cent had incomes of over 80 per cent of the expenditure. At the other extreme, 9.5 per cent of the households had no proper incomes at all. When the two areas are compared, the position in government colonies turns out to be worse than in other localities.

LOANS

2.21. The households obtained loans both from government and non-government sources. Taking the households of all localities together, 29.1 per cent had outstanding government loans and the average amount outstanding per household was Rs. 1025.3 per household. 44.6 per cent received non-government loans and the amount outstanding per household was Rs. 1289.3. Taking government and non-government loans together, 55.9 per cent of the households had loans and the average amount was Rs. 1567.2 per household. 44.1 per cent of households had no outstanding loans. The outstanding loans of all kinds per household was greater in other localities than in government colonies. The proportion of households having no outstanding loan was also greater in other localities (59.2 per cent) than in government colonies (36.0 per cent).

2.22. *Government loans :* These loans were given under different heads, namely house-building, trade, profession, maintenance loan, education loan, maintenance grant and education grant. 94.7 per cent of house-building loan was utilised in government colonies and the extent of utilisation of trade loans was only 75.1 per cent. The maintenance loan was fully utilised and only Rs. 101.6 was repaid by each borrowing household. The maintenance grant was only partly utilised, the extent of utilisation being 55.6 per cent. There was no education loan and education grant at all. Barring repayment of maintenance loan other loans were either repaid in almost negligible proportions or not at all.

In other localities, government loans were given under all the heads except house-building. Maintenance and education loans were fully utilised of which Rs. 105.0 and Rs. 600.0 respectively were repaid by each borrowing household. The extent of utilisation of trade, profession and maintenance grant was respectively 82.2, 85.5 and 70.9 per cent. The education grant was fully utilised. There was little or no repayment except under heads already mentioned.

All the households receiving trade and professional loans did not utilise them to the same extent. A little over 53 per cent of households receiving trade loans and nearly 45.0 per cent receiving professional loans utilised more than 90 per cent of these loans for these purposes. 20.6 per cent of households with trade loans and 27.6 per cent of households with professional loans could utilise their loans only up to 10 per cent. 13.5 per cent of households having trade loans and 13.8 per cent having professional loans could utilise more than 50 per cent of their loans, but not more than 70 per cent for these purposes.

HOUSING AND MISCELLANEOUS STATISTICS

2.23. *Housing :* In government colonies 98.9 per cent of households lived in rented accommodation and 1.1 per cent in owned accommodation. In other localities the position was somewhat different. 89.3 per cent of households were in rented houses and 10.7 per cent in houses owned by them. The proportion in all localities taken together was that 95.6 per cent of households lived in rented accommodation while 4.4 per cent of households occupied houses owned by them.

In government colonies more than 85 per cent of households had accommodation containing one room while in other localities only 41.0 per cent of households were in one-room accommodation. 10.3 per cent of households in government colonies and 37.2 per cent in other localities were living in two-room accommodation. The proportion of households having three-room accommodation in both government colonies and other localities were respectively 2.6 and 14.1 per cent. Many of the households shown as occupying one-room accommodation in government colonies had rooms in notional rather than in real sense. In such cases, the floor space occupied by a household was reckoned as one room.

The proportion of households occupying floor space up to 300 sq. ft. in government colonies and other localities was respectively 85.9 per cent and 72.0 per cent. The remaining households in both the areas occupied more than 300 sq. ft. of floor space. Taking all the localities together the average floor space per person was 46.7 sq. ft.

In all localities taken together 16.2 per cent of households were living in totally brick-built structures, 64.2 per cent in partly brick-built and partly non-brick-built structures and 19.6 per cent in totally non-brick-built ones.

2.24. *Lighting :* In government colonies, 37.1 per cent of households used electric light, 61.7 per cent used kerosene oil light and 1.2 per cent other types of light. In other localities, the proportion was 71.5 per cent of households for electric light and 28.5 for kerosene oil light.

2.25. *Water supply :* Over 80 per cent of the households got their drinking water from taps. In government colonies the proportion of such households was 82.9 per cent and in other localities it was 84.1 per cent. The remaining households got their drinking water from tube wells, wells and other sources. The average distance of the households from taps was 22.4 yards in government colonies and 11.5 yards in other localities.

In government colonies, only 7.2 per cent of households did not share the use of the source with other households. The proportion for other localities was 40.2 per cent. At the other extreme nearly 50 per cent of the households living in government colonies shared the use of the source with more than 30 households, the corresponding proportion of households in other localities being 14.6 per cent. If the proportion of households sharing the source with 10 others is considered, the percentage for government colonies comes to 86.5 against 27.8 for other localities.

2.26. *Latrines :* In government colonies 3.1 per cent of the households were using latrines with underground drainage and flush, 17.0 per cent latrines of the septic tank type and 78.4 per cent latrines of the old type. In other localities, 46.9 per cent were using latrines with underground drainage and flush, 1.3 per cent latrines of the septic tank type and 42.4 per cent the old type latrines. The average distance, taking all types together, was 30.2 yards in government colonies and 27.3 yards in other localities.

Only 4.4 per cent of households in government colonies were using latrines without sharing them with others, the corresponding proportion in other localities being 27.7 per cent. At the other end 49.2 per cent in government colonies shared latrines with more than 30 households whereas the percentage for other localities was 5.7. In government colonies 76.8 per cent of the households were using latrines with more than 10 others and the corresponding proportion for other localities was 22.2 per cent.

2.27. *Preference regarding means of livelihood :* Considering the households of all localities 61.9 per cent preferred activities which they were following at the time of the survey. The proportions of households preferring to remain in the same activities varied from 70.3 per cent in trade to 51.9 per cent in construction.

2.28. *Preference regarding place of habitation:* 71.9 per cent of all households of all localities preferred to live in the Bombay State and 22.6 per cent of households were indifferent as to the place of habitation. Only 5.5 per cent of the households preferred to go out of Bombay.

2.29. *Knowledge of local language :* Among all households of all localities 65.8 per cent could not read, write, speak or understand the local language. The proportions of this type of households were 75.4 per cent in government colonies and 48.0 per cent in other localities. It is, however, worth mentioning that the majority of these households had a working knowledge of Hindustani.

Chapter Three

DEMOGRAPHIC STATISTICS

3.1. The number of sample households in government colonies, in other localities and in all localities together with the corresponding number of persons and average size of households are given in Table (3.1) below.

TABLE (3.1) : NUMBER OF SAMPLE HOUSEHOLDS AND PERSONS PER HOUSEHOLD

areas	households		persons		persons per household
	number	percentage	number	percentage	
(1)	(2)	(3)	(4)	(5)	(6)
1. government colonies	880	65.2	4288	62.0	4.9
2. other localities	469	34.8	2626	38.0	5.6
3. all localities	1349	100.0	6914	100.0	5.1

3.2. The average size of households of the displaced persons was higher than the average size of households of the general population in the urban areas of the Bombay State which according to the 1951 Census was about 4.7 persons. The size of households living outside government colonies was of 5.6 persons, and of those living in government colonies of 4.9 persons. The smaller size in the government colonies might have been due to the tendency of the inmates to split themselves into smaller households to gain some advantage regarding housing space, ration facilities and government help.

3.3. A distribution of the sample households by their size as obtained in this survey is given in Table (3.2) below. Similar figures for the general population of the urban areas of the Bombay State as obtained in the 1951 Population Census are also given in the table. The 1951 Census figures for the displaced population on this distribution are not available.

TABLE (3.2) : DISTRIBUTION OF HOUSEHOLDS BY THEIR SIZE

size of households (persons)	sample survey : July–September 1953						1951 Census : urban general population (based on sample)	
	government colonies		other localities		all localities			
	number	percentage	number	percentage	number	percentage	number	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. 3 or less	294	33.4	120	25.6	414	30.7	34,481	38.5
2. 4—6	363	41.3	190	40.5	553	40.9	35,509	39.6
3. 7—9	186	21.1	114	24.3	300	22.3	14,201	15.8
4. 10 or above	37	4.2	45	9.6	82	6.1	5,473	6.1
5. total	880	100.0	469	100.0	1349	100.0	89,664	100.0
6. over-all average size	4.9		5.6		5.1		4.7	

3.4. Taking all localities together, the size of 28.4 per cent of the households was over 6 persons. The proportion for the general population living in the urban areas of the Bombay State was 21.9 per cent. Thus, there was a higher proportion of households with larger size amongst the displaced population as compared with the general population and this was particularly so amongst those who were living outside the government colonies.

DISTRIBUTION BY MOTHER TONGUE

3.5. The detailed figures regarding the distribution of displaced households by their mother tongue are given in Table (2) at the end of the report, but a summary is given in Table (3.3) below.

TABLE (3.3) : PERCENTAGE DISTRIBUTION OF HOUSEHOLDS BY MOTHER TONGUE

mother tongue	percentage distribution		
	government colonies	other localities	all localities
(1)	(2)	(3)	(4)
1. Sindhi	88.3	67.2	81.0
2. Punjabi	2.7	16.2	7.4
3. Gujrati	4.0	10.4	6.2
4. Marathi	1.5	1.3	1.4
5. others	2.5	4.5	3.2
6. no information	1.0	0.4	0.8
7. total	100.0	100.0	100.0
8. number of households	880	469	1349

3.6. Taking the households of all localities together, 81.0 per cent of the households were speaking Sindhi in their homes, 7.4 per cent were speaking Punjabi, 6.2 per cent Gujrati, 1.4 per cent Marathi and 3.2 per cent other languages. Information was not available from 11, i.e., 0.8 per cent of the households. It will be noticed that there was a greater concentration of Sindhis in government colonies. 88.3 per cent of the households in government colonies were Sindhis as compared to a proportion of 67.2 per cent for other localities. The proportion for Punjabis in the government colonies was 2.7 per cent as compared with 16.2 per cent in other localities. For Gujratis, the proportion in government colonies was 4.0 per cent as compared with 10.4 per cent in other localities. The proportions for the households speaking Marathi or other languages were low both in government colonies and in other localities.

3.7. Comparable figures in this regard are not available from 1951 Population Census, but some figures regarding the place of origin of migration of the displaced persons are available from a note written on Displaced Population by the Superintendent of Census, Bombay State, and these are given below.

TABLE (3.4) : PERCENTAGE DISTRIBUTION OF DISPLACED PERSONS OF BOMBAY BY STATES FROM WHICH THEY MIGRATED : 1951 POPULATION CENSUS

	place of origin	percentage distribution
	(1)	(2)
1.	Sind	82.0
2.	West Punjab	9.4
3.	Khairpur State	4.4
4.	others	4.2
5.	total	100.0

3.8. These figures relate to individuals for the State as a whole including both the rural and urban areas. Again, the figures refer to places of origin of migration and not specifically to mother tongue. Since the individuals migrating from the same State did not necessarily speak the same language, a distribution of persons arranged by place of origin can be only broadly compared with a distribution of households arranged by mother tongue. After taking these approximates into account, it will be seen that the figures of both sample survey and 1951 Census indicate that over four-fifths of the displaced persons are Sindhis. The remaining fifth, as observed in this survey, is composed of Punjabis, Gujratis and others.

AGE DISTRIBUTION

3.9. The distribution of the members of the sample households by age-group and sex has been given in Table (3). Only percentage distribution of members of households living in colonies and outside and similar distribution from the 1951 Population Census are given in the following table.

TABLE (3.5) : PERCENTAGE DISTRIBUTION OF DISPLACED PERSONS BY AGE GROUPS

	age group in years	sample survey : July–September : 1953			1951 Census	
					Bombay, Saurashtra and Kutch: urban and rural	
		government colonies	other localities	all localities	displaced	general
	(1)	(2)	(3)	(4)	(5)	(6)
1.	0	2.0	1.9	2.0	0.8	3.2
2.	1— 4	10.3	9.7	10.1	5.3	10.6
3.	5—14	27.1	27.9	27.4	28.1	25.7
4.	15—24	20.8	20.0	20.5	22.9	18.2
5.	25—34	14.2	14.4	14.3	17.4	15.9
6.	35—44	10.4	11.1	10.6	11.3	11.6
7.	45—54	6.9	7.6	7.1	7.5	7.7
8.	55—64	5.3	3.9	4.8	4.5	4.6
9.	65—74	2.7	3.1	2.9	1.6	1.8
10.	75 & above	0.3	0.4	0.3	0.6	0.7
11.	total	100.0	100.0	100.0	100.0	100.0

3.10. The figures of columns (5) and (6) relating to displaced and general population of urban and rural areas of Bombay, Saurashtra and Kutch have been obtained from the note on DPs written by the Superintendent of Census, Bombay State. Similar 1951 Census figures of displaced persons for urban areas of the Bombay State only were not readily available. The figures of the present survey, however, refer only to displaced population in urban areas of Bombay State for the period July to September 1953. Therefore, the figures from the two sources are only roughly comparable.

3.11. It will be seen from column (5) that there had been a gap in the age groups (0) and (1–4) years, the reason being that, in the census enumeration the children of the displaced persons who were born in India were not considered displaced. In the sample survey, on the other hand, the children of the displaced persons born in India were included as DPs. Consequently, the proportions in the first two age-groups, namely (0) and (1–4) years, are much higher and are near about the proportions for the general population. The age structures of the displaced population and of the general population are of the same order and as already noted by the Superintendent of Census of the Bombay State, the structure shows the signs of mass exodus.

LITERACY AND EDUCATION STANDARD

3.12. The distribution of persons by age group, sex and education standard living in different localities is given in Table (4). Only the percentages of literacy as obtained in this survey and in the 1951 Census are given below in Table (3.6).

TABLE (3.6): PERCENTAGE OF LITERACY (URBAN AREAS)

areas/population (1)	male (2)	female (3)	total (4)
Sample Survey : July–September 1953			
1. government colonies	65.8	37.1	51.5
2. other localities	74.3	56.7	65.7
3. all localities	69.1	44.5	56.9
1951 Census : Bombay State			
4. displaced population	69.7	38.7	57.6
5. general population	52.0	26.7	40.6

3.13. The proportion of literacy in the government colonies was 51.5 per cent as compared with 65.7 per cent for the other localities. A higher proportion of males and a still higher proportion of females living outside the colonies were literate. A comparison of the percentages of this survey with those of the 1951 Census for the displaced population shows a good agreement except that the percentage of literacy for the females is somewhat higher in the sample survey. As compared with the general urban population, the displaced population had a much higher proportion of literates.

3.14. A further breakdown is given in Table (3.7) showing the distribution of persons by education standard and by sex in the different localities.

TABLE (3.7) : PERCENTAGE DISTRIBUTION OF PERSONS BY EDUCATION STANDARD

standard	government colonies			other localities			all localities		
	male	female	total	male	female	total	male	female	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. illiterate	34.2	62.9	48.5	25.7	43.3	34.3	30.9	55.5	43.1
2. merely literate	44.8	30.7	37.8	38.7	39.8	39.3	42.5	34.1	38.4
3. under-matric	14.1	5.7	9.9	22.3	12.3	17.4	17.3	8.2	12.8
4. matric	5.9	0.7	3.3	10.5	3.7	7.1	7.6	1.8	4.7
5. graduate	1.0	0.0	0.5	2.8	0.9	1.9	1.7	0.4	1.0
6. total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

3.15. It will be seen that both in the case of males and females the people living outside the government colonies have better educational qualifications. Taking the males of all localities, 30.9 per cent were illiterate and 42.5 merely literate. 9.3 per cent passed the matriculation examination and of them, 1.7 per cent were graduates. Amongst the females of all localities, only 2.2 per cent passed the matriculation examination of whom only 0.4 per cent were graduates. Roughly, 90 per cent of the females were either illiterate or merely literate.

ECONOMIC STATUS

3.16. The distribution of persons by the economic status of earners, earning dependents, and non-earning dependents is given below in Table (3.8). Exactly comparable figures from the 1951 Census are not available because the note on displaced population prepared by the Census Superintendent of Bombay gives the figures for population of urban and rural areas taken together. To make the figures somewhat comparable only the figures relating to the non-agricultural population have been shown in this table.

TABLE (3.8) : PERCENTAGE DISTRIBUTION OF EARNERS, EARNING DEPENDENTS AND NON-EARNING DEPENDENTS

areas/population	percentage distribution			
	earners	earning dependents	non-earning dependents	total
(1)	(2)	(3)	(4)	(5)
Sample Survey : July–September 1953 (urban areas)				
1. government colonies	18.9	3.3	77.8	100.0
2. other localities	19.5	3.5	77.0	100.0
3. all localities	19.1	3.4	77.5	100.0
1951 Census : Bombay State (urban and rural, but of non-agricultural classes only)				
4. displaced population	26.3	2.7	71.0	100.0
5. general population	31.4	7.2	61.4	100.0

3.17. It will be seen that the proportions of earners, earning dependents and non-earning dependents in the colonies and outside were roughly the same but in comparison with the figures of the 1951 Census, the proportions are slightly different. The general picture for the displaced population, however, remained the same, namely, the ratio of earners and earning dependents to non-earning dependents remained 1 to 3 persons roughly compared with the ratio of 1 to 2 persons for the general population.

Distribution by livelihood classes

3.18. The detailed figures of the distribution of the economically active population, that is, earners and earning dependents, by livelihood classes and employment status are given in Table (7). A summary of these figures is given below in Table (3.9) together with the relevant figures of the 1951 Census. It may be noted that the figures of the sample survey show the distribution of only the economically active population. The census figures, on the other hand, show the distribution of earners including their dependents. If there is a pattern of variation in the size of dependents depending on the means of livelihood, then these two ways of distribution will give two sets of figures even for the same population. It may also be noted that the figures relate to two different years, namely 1951 and 1953.

3.19. The classification of the means of livelihood is also slightly different. For instance, under the classification "agriculture", only the activities depending on agricultural cultivation have been included in the census, whereas in the sample survey such activities as animal husbandry, fishing, forestry etc., have also been included in "agriculture". These differences, however, are not of much importance in this discussion because almost the whole of the active population covered in this survey was non-agricultural and the classification regarding non-agricultural activities is similar and comparable.

TABLE (3.9) : PERCENTAGE DISTRIBUTION OF ECONOMICALLY ACTIVE POPULATION BY MEANS OF LIVELIHOOD CLASSES

occupation	sample survey : July–September 1953			Census 1951 : Bombay State	
	government colonies	other localities	all localities	displaced population (urban and rural)	general population (urban)
(1)	(2)	(3)	(4)	(5)	(6)
1. agriculture	0.3	0.2	0.3	1.9	15.4
2. other production	16.8	15.2	16.2	13.0	28.9
3. trade	42.1	47.4	44.1	40.1	19.3
4. transport	4.9	5.3	5.1	4.1	5.2
5. other services and miscellaneous sources	35.9	31.9	34.3	40.9	31.2
6. total	100.0	100.0	100.0	100.0	100.0

3.20. The proportion occupied in agriculture was negligible both in government colonies and other localities. The proportions engaged in other kinds of production were more or less the same for the government colonies and other localities and also the proportions engaged in transport operations were roughly the same. The proportion engaged in trade was somewhat higher amongst the non-colony population and on the other hand the proportion engaged in other services was higher amongst the population living in government colonies.

3.21. When the proportions for all localities taken together are compared with those obtained in the 1951 Census, the figures are somewhat different. The census figures for the displaced population refer to both urban and rural areas and the figures for the urban areas only have not been available. However, the broad feature, namely, the concentration of displaced persons in trade and various services, come out clearly from both the sets of figures. It may be noted that "other services and miscellaneous sources" include personal services, construction work, hotel-keeping, repairing services, professional services and miscellaneous receipts from rent, charity etc. Again, when these figures are compared with those for the general urban population of the State, it is seen that while the displaced persons were concentrated mainly in commerce and other services, a much higher proportion of the general urban population was engaged in agriculture and other production.

3.22. The distribution of the economically active population by employment status is given below.

TABLE (3.10) : PERCENTAGE DISTRIBUTION OF ECONOMICALLY ACTIVE POPULATION BY EMPLOYMENT STATUS

employment status	government colonies	other localities	all localities
(1)	(2)	(3)	(4)
1. employer	0.7	2.8	1.5
2. employee	40.0	41.0	40.4
3. own account	57.8	54.9	56.7
4. unearned income	1.5	1.3	1.4
5. total	100.0	100.0	100.0

3.23. In the government colonies, about 98 per cent of the economically active persons were either employees or were working on own account. A very small proportion lived entirely on unearned incomes from sources such as rent, charities etc. In other localities, about 96 per cent also were either employees or working on own account. The proportion of employers, though very small, was significantly higher than that for government colonies.

Intensity of employment

3.24. The Table (3.11) below gives some idea of the intensity of employment and individual earnings in different means of livelihood.

TABLE (3.11) : DISTRIBUTION OF EARNERS AND EARNING DEPENDENTS (NON-EARNING DEPENDENTS EXCLUDED) BY LIVELIHOOD CLASSES WITH AVERAGE NUMBER OF DAYS EMPLOYED AND AVERAGE INCOME EARNED PER MONTH PER PERSON

		government colonies			other localities			all localities		
	livelihood classes	number of persons	days employed	income in Rs.	number of persons	days employed	income in Rs.	number of persons	days employed	income in Rs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1.	agriculture	3	24	36.50	1	—	300.00	4	18	102.38
2.	crafts and cottage industries	120	23	47.59	61	24	86.88	181	23	60.83
3.	large industries	40	24	100.56	31	23	141.60	71	24	118.48
4.	trade	39	25	63.43	287	25	116.00	686	25	85.42
5.	construction	27	20	96.77	13	22	187.27	40	21	126.18
6.	transport	47	25	107.13	32	25	132.03	79	25	117.22
7.	services	245	25	76.58	135	24	108.25	380	25	87.83
8.	professional services	59	22	95.54	36	23	157.29	95	22	118.82
9.	others	10	—	100.22	9	3	139.55	19	2	118.85
10.	total	950	24	71.78	605	24	118.14	1555	24	89.81

3.25. Taking all economically active persons of all livelihood classes in the government colonies, it will be seen that on average a person was employed for 24 days in a month during the period of the survey. The income per person for a month was about Rs.71.8. As compared with this, the average number of days employed for the other localities was also 24 days in a month but the average income per person was about Rs.118.1. The average position of the persons outside the government colonies was consistently better than that in the government colonies. When the income figures for the government colonies and the other localities are compared for each livelihood class, it will be seen that in all cases the income figures for the other localities are higher.

3.26. Information relating to income has been usually seen to be understated by the respondents. But since this is a general tendency, whatever be the value of the figures in the absolute sense, these are useful for relative comparisons.

That the average earnings of the economically active persons of the government colonies were lower than those for the other localities appears clear. Again, a comparison between the earnings in the different livelihood classes is interesting. It will be seen from the table that those who were engaged in large-scale industries had higher incomes than those engaged in crafts and cottage industries; those engaged in transport operations or construction jobs had higher incomes than those engaged in trades; those engaged in professional services had higher incomes than those engaged in services or in other types of odd jobs.

3.27. A further classification of the earners and earning dependents by the intensity of employment in terms of days in a month together with average incomes is given in Table (3.12) below.

TABLE (3.12) : DISTRIBUTION OF EARNERS AND EARNING DEPENDENTS (NON-EARNING DEPENDENTS EXCLUDED) BY DAYS OF EMPLOYMENT IN A MONTH WITH AVERAGE INCOME PER PERSON

days of employment in the month	government colonies			other localities			all localities		
	number of persons	percentage	income in Rs.	number of persons	percentage	income in Rs.	number of persons	percentage	income in Rs.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. 1—7	11	1.2	14.2	8	1.3	24.2	19	1.2	18.4
2. 8—14	28	2.9	30.7	6	1.0	46.1	34	2.2	33.4
3. 15—21	120	12.6	42.1	37	6.1	77.1	157	10.1	50.3
4. 22—30	791	83.3	78.5	554	91.6	123.0	1345	86.5	96.9
5. total	950	100.0	71.8	605	100.0	118.1	1555	100.0	89.8

3.28. Over four-fifths of the active population in the government colonies were working for more than 21 days in a month but as compared with the other localities a higher proportion of active persons of the government colonies was under-employed. The figures also show that the per person incomes increased as the number of days of employment increased, but here again, the average incomes in the government colonies were lower than the corresponding incomes of the other localities.

MEANS OF LIVELIHOOD BEFORE AND AFTER MIGRATION

3.29. The detailed figures showing the distribution of households by livelihood classes before migration and during the period of survey are given in Table (10). A summary of the details has been given in Table (3.13) below. It may be noted that the information is based on the principal means of livelihood of the principal earners of the households and is thus different from the figures showing the distribution of individual earners and earning dependents by livelihood classes. The figures in the table have been so arranged as to show in what proportion the principal earners of the households in the different livelihood classes still follow the activities they used to do before migration.

TABLE (3.13) : DISTRIBUTION OF HOUSEHOLDS BY PRE-MIGRATION LIVELIHOOD CLASSES AND PROPORTIONS CONTINUING THE SAME ACTIVITIES DURING THE PERIOD OF THE SURVEY

livelihood classes	government colonies			other localities			all localities		
	number of households		percentage continuing	number of households		percentage continuing	number of households		percentage continuing
	before migration	continuing now		before migration	continuing now		before migration	continuing now	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. agriculture	139	1	0.7	28	—	—	167	1	0.6
2. crafts and cottage industries	76	41	53.9	40	28	70.0	116	69	59.5
3. large industries	24	8	33.3	16	7	43.8	40	15	37.5
4. trade	333	205	61.6	215	147	68.4	548	352	64.2
5. construction	24	13	54.2	13	6	46.2	37	19	51.4
6. transport	17	9	52.9	11	5	45.5	28	14	50.0
7. services	162	83	51.2	88	45	51.1	250	128	51.2
8. professional services	56	38	67.9	32	21	65.6	88	59	67.0
9. others	37	11	29.7	20	7	35.0	57	18	31.6
10. total	868	409	47.1	463	266	57.5	1331	675	50.7
11. no information	12	—	—	6	—	—	18	—	—

3.30. In government colonies, taking all livelihood classes together, 47.1 per cent of the households continued to follow the same activities as they used to do before migration. In other localities, 57.5 per cent of the households continued to follow the old activities. Taking all the localities together, 50.7 per cent of the households continued to follow their old vocations.

3.31. The proportions for the different livelihood classes, however, vary. In the government colonies, only 1 of the 139 households (0.7 per cent) depending on agriculture continued to depend on the same. Such a large shift in the agricultural class would naturally occur because the households came away from their lands and were living after migration in urban areas. Because of the circumstances of migration even those who were non-cultivating owners would not get any rent from land. Only about a third of the households engaged in large industries or other types of odd jobs before migration continued to do so at present. 67.9 per cent of the households engaged before migration in professional services and 61.6 per cent of those engaged in trade continued their old activities in India. Similar proportions for crafts and cottage industries, construction, transport and services were a little above 50 per cent.

3.32. In other localities, proportions continuing the old vocations were generally higher. Only about a third of the households depending on various odd jobs before migration continued to do so. Roughly, a half of the households engaged in large scale industries, construction, transport and services before migration continued in the old lines. About two-thirds of the households engaged before migration in crafts and cottage industries, trade and professional services continued in their old lines.

3.33. Taking all the households of all localities together, roughly two-thirds of the households continued to follow the old vocations in crafts and cottage industries, trade and professional services. Roughly, half of the households continued the old vocations in construction, transport and services; and about a third continued in large industries and in other types of odd jobs. Practically all who before migration were depending on agriculture were no longer doing so.

3.34. Information regarding means of livelihood only was collected in this survey to compare the positions obtaining before and after migration. No attempt was made to collect similar information on income, expenditure or value of properties simply because such questions would have invited biased replies. Considering the emotional stress in which the displaced persons live, if they were asked as to what they were earning or spending or owning before they migrated as compared with their present position, it was highly likely that they would have given an optimistic picture of the good old days. There would be no means of checking the correctness of their statements and the figures would have been merely misleading.

CHAPTER FOUR

EXPENDITURE ON CONSUMPTION

4.1. The expenditure per household per month for the government colonies, other localities and for all localities taken together are given below in Table (4.1). For purposes of comparison, per household expenditures of the general population of some urban areas of the Bombay State have also been given. These figures were collected in the 4th round of the National Sample Survey in the period April–September 1952. The survey covered a sample of 5 towns in the State and these were Ahmedabad, Amalner, Malegaon, Manglawedha and Shigali. In addition, Bombay city was also covered.

TABLE (4.1): COMPARATIVE CONSUMER EXPENDITURE OF DISPLACED AND GENERAL POPULATION IN RUPEES PER HOUSEHOLD PER MONTH

areas	number of sample households	expenditure in Rs.
(1)	(2)	(3)
Sample Survey of displaced persons : July-September 1953		
1. government colonies	880	112.3
2. other localities	469	201.3
3. all localities	1349	143.2
National Sample Survey : April-September 1952		
4. 5 towns of Bombay	304	143.5
5. Bombay city	157	276.3

4.2. The per household expenditure of the displaced population of all urban localities taken together was Rs.143.2 per month in the period July–September 1953. The per household expenditure for the general population in the 5 towns of Bombay was practically the same, namely, Rs.143.5 per month, but for a different period, April–September 1952. The expenditure per household in the government colonies was low, e.g., Rs.112.3, in comparison with the figure for the other localities which was Rs.201.3. In turn, the figure for the non-colony displaced population is lower than the per household expenditure of the general population in Bombay city which was Rs.276.3 per month during April–September 1952.

4.3. Although the average expenditures of the displaced urban households and of the general urban households seem to be equal, there are two reasons for which the position of the displaced households turns out to be somewhat worse. Firstly, the average size of a displaced household was a little higher than a general household and secondly, there was a slight increase in living costs from the period of the 1952 survey to that of the 1953 survey.

4.4. It will be seen that the gap between the two periods of survey is about a year; and prices changed in course of that year. The Labour Bureau, Ministry of Labour, Government of India, publishes monthly in the Indian Labour Gazette cost of living index numbers for 4 towns of Bombay State, namely, Bombay, Ahmedabad, Sholapur and Jalgaon and the average index numbers for the periods April–September 1952 and July–September 1953 are given in Table (4.2) below.

TABLE (4.2): AVERAGE OF WORKING CLASS COST OF LIVING INDEX NUMBERS FOR THE TWO PERIODS, APRIL-SEPTEMBER 1952 AND JULY-SEPTEMBER 1953

(1939=100)

towns	April–September 1952	July–September 1953	increase in points	percentage increase
(1)	(2)	(3)	(4)	(5)
1. Bombay	323	355	32	10.0
2. Ahmedabad	363	391	28	7.7
3. Sholapur	424	448	24	5.7
4. Jalgaon	440	492	52	11.8

4.5. The increase in living costs between the two periods varied between 5.7 and 11.8 per cent and the overall order of increase was about, say, 10 per cent. In view of this increase in living costs and of larger household size, it will be correct to say that the level of living of the displaced urban population was somewhat lower than that for the general urban population of the State. Roughly, the level of living in the government colonies compares unfavourably with that of the general urban population but the level of living of the displaced population outside the colonies compares favourably with the general level, although, it is lower than the level obtained in Bombay city.

4.6. A comparison on the basis of average expenditures only does not give a satisfactory picture because the averages may be vitiated by extreme values of expenditures of very rich households or very poor households. Therefore, the distributions of households by different levels of monthly expenditure for these areas have been given in Table (4.3) below.

TABLE (4.3) : PERCENTAGE DISTRIBUTION OF DISPLACED HOUSEHOLDS AND GENERAL HOUSEHOLDS IN BOMBAY BY LEVELS OF MONTHLY EXPENDITURE

	displaced households : July–September 1953			general households : Bombay April–September 1952	
monthly expenditure in Rs.	government colonies	other localities	all localities	5 towns	city
(1)	(2)	(3)	(4)	(5)	(6)
1. up to 50	17.5	4.3	12.9	12.8	3.8
2. 51—100	40.6	16.2	32.1	32.3	19.1
3. 101—150	23.6	26.4	24.6	20.7	16.6
4. 151—300	14.9	37.3	22.7	26.3	32.5
5. 301—500	2.5	10.9	5.4	5.6	14.6
6. above 500	0.9	4.9	2.3	2.3	13.4
7. total	100.0	100.0	100.0	100.0	100.0
8. number of households	880	469	1349	304	157

4.7. These percentage distributions confirm that the living conditions (in terms of consumption expenditure) of the displaced households in the urban areas of the State taken as a whole were near but lower than the conditions of the general urban households. The figures, again, show that the conditions of the households living in government colonies were worse than the general urban households and that the conditions of the displaced households not living in government colonies were better than the general urban households. These non-colony households were concentrated in middle levels, e.g., 48.2 per cent between monthly expenditures of Rs.151 and Rs.500; and 74.6 per cent between expenditures of Rs.101 and Rs.500.

4.8. The comparative patterns of expenditure of displaced households living in government colonies and in other localities as also the patterns for the general population in the 5 sample towns and in Bombay city are given in Table (4.4) below.

TABLE (4.4) : PATTERNS OF CONSUMPTION OF DISPLACED HOUSEHOLDS IN DIFFERENT LOCALITIES AND OF GENERAL HOUSEHOLDS IN URBAN AREAS OF BOMBAY

	displaced households : July–September 1953						general households : Bombay April-September 1952	
	government colonies		other localities		all localities		5 towns	city
items of expenditure	expenditure in Rs.	percentage	expenditure in Rs.	percentage	expenditure in Rs.	percentage	percentage	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. food	71.4	63.7	110.3	54.8	84.9	59.4	48.2	37.7
2. clothing	9.7	8.6	19.0	9.4	12.9	9.0	10.5	10.1
3. fuel and light	7.1	6.3	11.2	5.6	8.5	5.9	7.2	5.2
4. rent	3.4	3.0	19.7	9.8	9.1	6.3	3.1	8.2
5. residual	20.7	18.4	41.1	20.4	27.8	19.4	31.0	38.8
6. total	112.3	100.0	201.3	100.0	143.2	100.0	100.0	100.0
7. number of households	880		469		1349		304	157

4.9. Poorer conditions of the households living in government colonies have been reflected in the pattern of their expenditure. A higher proportion of expenditure went for food, namely 63.7 per cent, as compared with 54.8 per cent for the households living in other localities. Again, the proportion for the residual items was lower for the government colonies. One very interesting point should be noted, namely, that the per household expenditure for house rent in government colonies was only Rs.3.4 as compared with Rs.19.7 per household in the other localities. A part of this low expenditure in government colonies might be due to lower rents but most of it was due to the fact that many households were not paying their dues.

4.10. Taking the households of all localities together, it will be seen that 59.4 per cent of the expenditure were spent for food and 19.4 per cent on miscellaneous items. As compared with this, 48.2 per cent of expenditure of general households in the 5 sample towns went for food and 31.0 per cent for miscellaneous items. A higher proportion of expenditure for food and a lower proportion for residual items indicate lower living standards and accordingly, it is clear that the living standards of the displaced households were lower than those of the general households in the urban areas. The contrast is sharper when the figures of the displaced households are compared with the pattern of the general households living in Bombay city.

Chapter Five

INCOME

5.1. Understatements of income and overstatements of debt by households have often been noted in economic surveys, and hence the figures regarding incomes or loans of this survey also should be read cautiously. One approach might have been not to collect any information regarding these aspects at all. But this would have left a gap which one would be tempted to fill with scrappy particulars from other sources even more unreliable. Therefore, provision was made for collecting data on income and loan, and it was felt that if the particulars regarding the receipts were recorded under a number of different items, an idea of at least the relative importance of these various sources could be obtained clearly.

5.2. The households were asked about their receipts and expenditures for a month preceding the date of interview and the figures obtained are given in Table (5.1) below.

TABLE (5.1) : AVERAGE MONTHLY RECEIPT AND EXPENDITURE PER HOUSEHOLD IN RUPEES

receipt/expenditure	government colonies	other localities	all localities
(1)	(2)	(3)	(4)
1. monthly receipt	123.1	225.0	158.5
2. monthly expenditure	112.3	201.3	143.2
3. number of households	880	469	1349

5.3. The monthly receipts were higher than the expenditures in both government colonies and other localities. It should, however, be noted that although the receipts through the various sources were meant for consumption, all of them were not incomes in proper sense. The households had to live on and their demands were met by money from government doles, from private help or loans or even by selling household assets such as ornaments.

5.4. Particulars regarding current receipts by the households for purposes of consumption were collected under 6 items, namely, (1) income from occupation, (2) receipts from other sources, (3) imputed value of supplies obtained from one's own enterprise, (4) government dole, (5) non-government help and (6) non-government loan.

5.5. Occupation meant productive activities involving direct participation. Rent receivers of all kinds and money lenders were excluded from this group and their incomes were entered as receipts from other sources. Although it was intended that incomes from subsidiary occupations would be recorded as incomes from occupation, in practice it turned out that most of such incomes were shown as receipts from other sources. The households were hesitant to declare all their subsidiary occupations.

5.6. Receipts from other sources therefore included incomes from most of the subsidiary activities and incomes from property of any kind and from various unproductive activities. These also included savings or balances left over from earlier months which were different from balances of loans. Receipts by loans taken for consumption purposes before the reference period of the investigation were also recorded as loans.

5.7. Imputed values of supplies obtained from one's own enterprise were taken into account because these supplies were included in the expenditures on consumption. As practically all the households covered in the survey were non-agriculturists, this was an unimportant item in the budget.

5.8. Help received from government for consumption purposes was recorded as "government dole" and help from all other sources as non-government help. Charities of all kinds and also remittances from friends and relatives were included in non-government help. Perhaps some of these remittances can be considered as proper in our present social set-up, such as, the remittance of a son to his parents and so on but no distinction was made in recording the different types of remittances.

5.9. Private loans taken only for consumption purposes were recorded as non-government loans. When loans were taken for enterprise purposes, they were excluded from consumption account and recorded separately in another part of the questionnaire.

5.10. The average monthly receipts over all sample households by these 6 breakdowns are shown below in Table (5.2).

TABLE (5.2) : DISTRIBUTION OF MONTHLY RECEIPTS PER HOUSEHOLD BY SOURCES OF RECEIPTS

source of receipt	government colonies		other localities		all localities	
	rupees per house-hold	percent-age	rupees per house-hold	percent-age	rupees per house-hold	percent-age
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. occupation	78.6	63.9	159.9	71.1	106.9	67.5
2. other sources	20.6	16.7	43.1	19.2	28.4	17.9
3. own supply	0.4	0.3	1.2	0.5	0.7	0.4
4. government dole	3.8	3.1	1.4	0.6	2.9	1.8
5. non-government help	4.7	3.8	11.0	4.9	6.9	4.4
6. non-government loan	15.0	12.2	8.4	3.7	12.7	8.0
7. total receipt	123.1	100.0	225.0	100.0	158.5	100.0
8. number of households	880		469		1349	

5.11. Taking the households of all localities together, the total receipt per household was Rs.158.5 of which 67.5 per cent were income from occupation and 17.9 per cent income from other sources. The imputed value of own supply was only 0.4 per cent. The proportion for government dole was 1.8 per cent and that for non-government help was 4.4 per cent. Loan taken from non-government sources for consumption purpose was 8.0 per cent.

5.12. The relative importance of the different sources varied somewhat between the government colonies and other localities and, on the whole, the position of the government colonies was worse. The total receipt per household in the government colonies was Rs.123.1, of which 63 9 per cent came as income from occupation and 16.7 per cent as income from other sources. The imputed value of own supply was 0.3 per cent. Government dole represented 3.1 per cent and non-government help 3.8 per cent. 12.2 per cent came by loan from non-government sources taken for consumption purposes.

5.13. The total receipt per household in other localities was Rs.225.0, of which 71.1 per cent were income from occupation and 19.2 per cent income from other sources. The imputed value of own supply was 0.5 per cent. The proportion for government dole was only 0.6 per cent. 4.9 per cent of total receipt came from non-government help and 3.7 per cent from loan taken for consumption purposes from non-government sources.

5.14. It may be noted that the above figures show the average position over all the households. All the households did not obtain money from all these six sources, but many did receive money from more than one source. An idea of the proportions of the households who had partly or wholly depended on the different sources may be obtained by showing the members of the households who had reported such receipts and the figures have been given in Table (5.3) below. If a household had obtained money from, say, 3 sources, it was then included in the numbers shown against all those 3 sources. When all the households receiving money from items 1 to 4 of the table have been accounted for, a balance of households will remain who must have obtained their money as incomes from occupations or from other sources.

5.15. Another point may also be noted. Money obtained by the sale of assets, such as ornaments, has been included in the item, income from other sources, in the preceding Table (5.2). These receipts have been separated in the table below

TABLE (5.3) : NUMBER AND PERCENTAGE OF HOUSEHOLDS DEPENDING PARTLY OR WHOLLY ON DIFFERENT SOURCES OF RECEIPTS FOR MEETING THEIR CONSUMPTION DEMANDS

source of receipt	government colonies		other localities		all localities	
	number of households	percentage of total	number of households	percentage of total	number of households	percentage of total
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. government dole	90	10.2	11	2.3	101	7.5
2. non-government help	79	9.0	50	10.7	129	9.6
3. non-government loan	210	23.9	43	9.2	253	18.8
4. sale of assets	48	5.5	13	2.8	61	4.5
5. residual	504	57.3	360	76.8	864	64.0
6. total	(880 =	100%)	(469 =	100%)	(1349 =	100%)

and shown as a different item so that receipts from other sources may represent proper incomes. Selling of assets was reported from 61 of the total households.

5.16. Taking households of all the localities together, 7.5 per cent received government doles, and 9.6 per cent received non-government help. 18.7 per cent of the households got loans from non-government sources for consumption purposes and 4.5 per cent got money by sale of assets. The residual proportion e.g., 59.7 per cent depended on their incomes alone.

5.17. In the government colonies, 10.2 per cent of households received government doles and 9.0 per cent non-government help. 23.9 per cent got loans from non-government sources for consumption purposes, and 5.4 per cent of the households got some money by selling assets. The remaining households, namely, 51.5 per cent depended on their incomes only.

5.18. In the other localities, 2.3 per cent of the households received government doles and 10.7 per cent non-government help. 9.2 per cent got loans for consumption purposes from non-government sources and 2.8 per cent got some money by selling assets. The remaining households, e.g., 75.0 per cent depended on their income only.

5.19. The amounts of money received per household through the four sources, which are not incomes but only receipts to cover deficits, are given in Table (5.4) below.

TABLE (5.4) : NUMBER OF HOUSEHOLDS RECEIVING AND THE AMOUNT RECEIVED PER HOUSEHOLD BY TYPE OF SOURCE

type of source	government colonies		other localities		all localities	
	number of households	receipt per household (in Rs.)	number of households	receipt per household (in Rs.)	number of households	receipt per household (in Rs.)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. government doles	90	36.7	11	61.1	101	39.4
2. non-government help	79	51.8	50	103.1	129	71.7
3. non-government loan	210	63.0	43	91.5	253	67.8
4. sale of assets	48	68.4	13	105.1	61	76.2

5.20. For all localities taken together, the average amount of government dole per receiving household was Rs.39.4. In government colonies, the amount of dole per receiving household was Rs.36.7 and in other localities the amount was Rs.61.1. The average amount of non-government help per receiving household was Rs.71.7 for the households of all localities taken together and Rs.51.8 for government colonies and Rs.103.1 for the other localities. Similar amounts of non-government loans were Rs.67.8 for all localities, Rs.63.0 for government colonies and Rs.91.5 for other localities. The amounts obtained per household by selling assets were Rs.76.2 for all localities, Rs.68.4 for government colonies and Rs.105.1 for other localities.

5.21. In view of the fact that 18.7 per cent of the total sample households reported that they got loans from non-government sources for purposes of consumption, a closer study of the nature of these loans has been made. The households which reported that they got loans for consumption have been arranged by the nature of the securities given by them and whether they had to pay any interest or not. The details in this regard have been given in Table (16) at the end of the report. A summary of these figures is given in Table (5.5) below.

TABLE (5.5) : NUMBER OF HOUSEHOLDS TAKING LOANS FOR CONSUMPTION WITH OR WITHOUT INTEREST BY TYPES OF SECURITIES

security	government colonies			other localities			all localities		
	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. none	5	46	51	2	24	26	7	70	77
2. personal	19	115	134	1	13	14	20	128	148
3. gold or others	4	10	14	1	1	2	5	11	16
4. not returned	1	10	11	–	1	1	1	11	12
5. total	29	181	210	4	39	43	33	220	253

5.22. In government colonies, of the 210 households who reported to have taken loans to meet consumption demands, only 29 stated that they had to pay interest for the loans and another 10 did not pay interest but got the loans against gold and other articles as securities. The rest, i.e., 171 households or 81.4 per cent stated that they got the loans without interest and without any proper security. This seems to be surprisingly high proportion for those who could get loans in such favourable terms. It will be seen that their proportion to total sample households for the government colonies (171 out of 880) is 19.4 per cent. In other words, roughly every fifth household in the government colonies could get loans from non-government sources without interest and without proper security.

5.23. The picture seems somewhat unrealistic and probably there had been incorrect statements about the loans and it is highly likely that a part of these households in understating their incomes had to explain the gap between expenditure and income by loans. From Table (5.3) of this chapter it will be seen that 23.9 per cent of the households in government colonies got loans from non-government sources and only 4.5 per cent of these got loans against interest or some securities. The statements of some of the remaining 19.4 per cent of the households are open to question.

5.24. In other localities, 43 households reported that they got such loans and 4 of them stated that they had to pay interest and one stated that it got loan without interest but against gold or other articles as securities. The remaining 38 or 88.4 per cent stated that they got their loans without interest and without proper securities. The proportion of these 38 households to the total sample of 469 is 8.1 per cent. This percentage is not so high but even here the possibilities of incorrect statements may not be ruled out altogether.

5.25. Taking the households of all localities together 253 reported that they got loans for consumption but 209 of these households or 82.6 per cent stated that the loans were obtained without any security or interest. Of the total sample households of 1349, 18.7 per cent were reported to have taken loans but, of this proportion, 15.5 per cent stated to have taken loans against no securities or interest and probably some of these statements in respect of borrowing are not correct.

5.26. To what extent the incomes of the households could meet the expenditures was also examined. In doing so, only the receipts which can be considered as proper incomes have been taken into consideration, and receipts from government dole, non-government help, non-government loan and from sale of assets have been excluded. It may also be noted that no adjustments have been made for possible understatements regarding incomes. For each household, its income was matched against its expenditure and the former has been expressed as percentage of the latter. For instance, if the income of a household was only half of its expenditure, the percentage would be 50; or if its income just covered its expenditure, the percentage would be 100, and so on. The households have then been arranged by the levels of percentages of income to expenditure and the figures have been given in Table (5.6) below.

TABLE (5.6) : DISTRIBUTION OF HOUSEHOLDS BY INCOME TO EXPENDITURE PERCENTAGE LEVELS

percentage income to expenditure	government colonies			other localities			all localities		
	number of households	percentage	cumulative percentage	number of households	percentage	cumulative percentage	number of households	percentage	cumulative percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. nil	109	12.4	100.0	19	4.1	100.0	128	9.5	100.0
2. up to 20	24	2.7	87.6	8	1.7	95.9	32	2.4	90.5
3. 21— 40	40	4.6	84.9	14	3.0	94.2	54	4.0	88.1
4. 41— 60	72	8.2	80.3	23	4.9	91.2	95	7.0	84.1
5. 61— 80	91	10.3	72.1	38	8.1	86.3	129	9.6	77.1
6. 81— 90	42	4.8	61.8	15	3.2	78.2	57	4.2	67.5
7. 91—100	183	20.8	57.0	134	28.6	75.0	317	23.5	63.3
8. above 100	319	36.2	36.2	218	46.4	46.4	537	39.8	39.8
9. total	880	100.0	—	469	100.0	—	1349	100.0	—

5.27. Taking the households living in all the localities together, 39.8 per cent of the households had incomes above expenditure, 63.3 per cent had incomes over 90 per cent of their expenditure and 67.5 per cent had incomes over 80 per cent of the expenditure. At the other extreme, 9.5 per cent of the households had no proper incomes at all and were living entirely on government doles, non-government help, non-government loans or by sale of assets.

5.28. In the government colonies, 36.2 per cent of the households had incomes larger than their expenditures, 57.0 per cent had incomes over 90 per cent of their expenditures and 61.8 per cent over 80 per cent of their expenditures. At the other extreme, 12.4 per cent of the households had no proper incomes at all and were depending on various kinds of help.

5.29. In other localities 46.4 per cent of the households had incomes larger than their expenditures and 75.0 per cent had incomes over 90 per cent of their expenditures, and 78.2 per cent had incomes over 80 per cent of their expenditures. At the other extreme, 4.1 per cent ofthe households had no proper incomes and were depending on various kinds of help.

5.30. The relative positions of the households living in government colonies and in other localities are clear from the figures, namely, that a higher proportion of households living outside the colonies is now on their feet as compared with those living in the colonies. But in considering the absolute position regarding the proportion earning enough to cover their expenses, one should be cautious in interpreting the figures of Table (5.6). As discussed in paragraphs 5.22–5.25, there are indications that incomes were probably understated, although it is not possible to indicate the extent of understatement. Even though unsatisfactory, some arbitrary level may be adopted to allow for understatement of incomes in order that one may get at some concrete figures. For instance, if it is assumed that all households who by their incomes covered over 80 per cent of their expenses be considered in effect as covering whole expenses, then 61.8 per cent of the households in the government colonies and 78.2 per cent of those living outside the government colonies were self-supporting at the time of this survey. But whether that level (ratio of income to expenditure) should be taken at 90 per cent or, 80 per cent or, 60 per cent will depend on the judgment of those who are familiar with the local conditions.

CHAPTER SIX

LOANS

6.1. The households obtained loans from the Government and also from non-government sources. Particulars regarding outstanding loans (for consumption and enterprise) were collected in the survey and the figures showing the households having outstanding government loans, non-government loans and both government and non-government loans have been given below in Table (6.1) along with average amount of outstanding loans per household.

TABLE (6.1) : NUMBER OF HOUSEHOLDS WITH AN AMOUNT PER HOUSEHOLD OF OUT-STANDING LOANS BY TYPES OF LOANS

	government colonies			other localities			all localities		
type of loans	number of households	percentage of total	average amount (in Rs.)	number of households	percentage of total	average amount (in Rs.)	number of households	percentage of total	average amount (in Rs.)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. government loans	321	36.4	806.7	73	15.6	1986.8	394	29.1	1025.3
2. non-government loans	438	49.8	866.8	164	35.0	2417.7	602	44.6	1289.3
3. government cum non-government loans	562	64.0	1136.3	191	40.8	2835.3	753	55.9	1567.2
4. no loans	318	36.0	—	278	59.2	—	596	44.1	—
5. number of households	880			469			1349		

6.2. Taking the households of all localities together, 29.1 per cent of the households had outstanding government loans and the average amount outstanding was Rs.1025.3 per household. 44.6 per cent of the households had non-government loans and the amount outstanding per household was Rs.1289.3. Taking government and non-government loans together, 55.9 per cent of the households had loans and the average amount was Rs.1567.2 per household. 44.1 per cent of the households had no outstanding loans.

6.3. In the government colonies, 36.4 per cent of the households had government loans and the average amount was Rs.806.7 per household. 49.8 per cent of the households had non-government loans and the average amount per household was Rs.866.8. Taking government and non-government loans together, 64.0 per cent of the households had loans and the amount was Rs.1136.3 per household. 36.0 per cent of the households had no outstanding loans.

6.4. In other localities, 15.6 per cent of the households had government loans but the size of loans was larger and the average amount was Rs.1986.8 per household. 35.0 per cent of households had non-government loans and the per household amount was Rs.2417.7. Taking government and non-government loans together, 40.8 per cent of the households had loans and the average amount was Rs.2835.3 per household. 59.2 per cent of the households had no outstanding loans.

6.5. A distribution of the households by the size of the outstanding government and non-government loans taken together is given below.

TABLE (6.2) : DISTRIBUTION OF HOUSEHOLDS BY SIZE OF OUTSTANDING GOVERNMENT AND NON-GOVERNMENT LOANS

		government colonies		other localities		all localities	
	level of outstanding loans (Rs.)	number of households	percentage	number of households	percentage	number of households	percentage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	0	318	36.0	278	59.2	596	44.1
2.	up to 499	185	21.2	50	10.7	235	17.5
3.	500 — 999	182	20.7	44	9.4	226	16.7
4.	1000 — 2999	158	18.0	57	12.2	215	16.0
5.	3000 — 4999	17	1.9	15	3.2	32	2.4
6.	5000 — 9999	18	2.0	12	2.6	30	2.2
7.	10000—14999	1	0.1	8	1.7	9	0.7
8.	15000—19999	—	—	3	0.6	3	0.2
9.	20000 & above	1	0.1	2	0.4	3	0.2
10.	total	880	100.0	469	100.0	1349	100.0

6.6. Taking the households of all localities together, 17.5 per cent had loans below Rs.500, and 34.2 per cent below Rs.1000. Only 5.7 per cent of households had loans amounting to Rs.3000 and above.

6.7. In the government colonies, 21.2 per cent of the households had loans below Rs.500, and 41.9 per cent below Rs.1000. 4.1 per cent of the households had loans amounting to Rs.3000 or above. There were only two households who had outstanding loans over Rs.10,000.

6.8. In other localities, 10.7 per cent of the households had loans below Rs.500, and 20.1 per cent below Rs.1000. 8.5 per cent of the households had loans amounting to Rs.3000 or above. There were as many as 13 households (2.7 per cent) whose size of outstanding loans was over Rs.10,000.

GOVERNMENT LOANS

6.9. Particulars regarding government loans were obtained in greater details and not only information on outstanding loans but also the details of these loans from the beginning were obtained. Full particulars in this regard have been given in Tables (21) to (24) at the end, but some of the more important figures have been given in the summary tables of this chapter. Government loans represent loans from all Government Agencies including the Rehabilitation Finance Administration.

TABLE (6.3) : RECEIPTS, UTILISATION AND REPAYMENT PER HOUSEHOLD OF GOVERNMENT LOANS AND GRANTS FOR DIFFERENT PURPOSES IN GOVERNMENT COLONIES

type of loans and grants	number of households=880					
	number of households	amount received (in Rs.)	extent utilised		amount repaid up to the period of survey (in Rs.)	interest paid up to the period of survey (in Rs.)
			amount (Rs.)	percentage		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. house-building	3	4750.0	4500.0	94.7	—	—
2. trade	284	796.4	598.2	75.1	8.7	4.6
3. profession	25	619.0	454.6	73.4	0.2	0.0
4. maintenance loan	12	563.3	563.3	100.0	101.6	0.0
5. education loan	—	—	—	—	—	—
6. maintenance grant	4	787.5	437.5	55.6	—	—
7. education grant	—	—	—	—	—	—

6.10. A small number of households received more than one type of loan and hence, it may be mentioned, that the total of households shown in column (2) will be larger than the total of households in government colonies who received loans. The majority of loans were trade and professional loans, and trade loans included loans to craftsmen also. 284 of the households received trade loans and the average amount came to Rs.796.4 per household. As stated by these households, 75.1 per cent of the amount were, on the whole, utilised for setting up their trades. The remaining part was utilised for consumption, repayment of private loans and similar other purposes. 25 of the households took professional loans and the amount per household was Rs.619.0. On average 73.4 per cent of the amount were utilised for setting up their professions. 12 of the households obtained maintenance loans and the amount per household was Rs.563.3. It may be noted that an average amount of Rs.101.6 per household was paid back to the Government. 4 of the households obtained maintenance grants and the amount per household was Rs.787.5 but, on average, only 55.6 per cent were utilised for maintenance and the rest utilised for repaying private loans. 3 of the households got house-building loans and the amount per household was Rs.4750.0, 94.7 per cent of which had so far been utilised for house-building and the rest remained as cash balance.

6.11. The particulars in regard to households living in other localities are given in Table (6.4) below.

TABLE (6.4) : RECEIPTS, UTILISATION AND REPAYMENT PER HOUSEHOLD OF GOVERNMENT LOANS AND GRANTS FOR DIFFERENT PURPOSES IN OTHER LOCALITIES

	number of households=469					
type of loans and grants	number of households	amount received (in Rs.)	extent utilised		amount repaid up to the period of survey (in Rs.)	interest paid up to the period of survey (in Rs.)
			amount (Rs.)	percentage		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. house-building	—	—	—	—	—	—
2. trade	66	2117.4	1741.4	82.2	4.5	18.8
3. profession	4	950.0	812.5	85.5	—	—
4. maintenance loan	2	550.0	550.0	100.0	105.0	—
5. education loan	1	1500.0	1500.0	100.0	600.0	0.0
6. maintenance grant	4	123.5	87.5	70.9	—	—
7. education grant	2	133.0	133.0	100.0	—	—

6.12. 66 of the households living in other localities received trade loans but the amount per household was much larger than that for the government colonies, namely, Rs.2117.4. On average, 82.2 per cent of this amount were utilised for setting up their trade and more than half of the remainder was still kept in hand as balance. 4 of the households got professional loans and the amount was Rs. 950 per household. 85.5 per cent of this amount were utilised for setting up their professions and a good part of the rest was used for consumption. Only 2 households got maintenance loans and the amount per household was Rs.550.0, the whole of which was utilised for consumption. Here again, it is interesting to note that on average Rs.105.0 were paid back by these two households. One household received an education loan of Rs.1500 and by the time this survey was carried out, Rs.600 of this amount were paid back. 4 of the households obtained maintenance grant and the amount per household was Rs.123.5 of which 70.9 per cent were utilised for maintenance and the rest was held in hand as balance. 2 of the households were given education grants and the amount per household was Rs.133.0 which was utilised entirely for education.

6.13. The figures for the households of all localities taken together are given in Table (6.5) below.

TABLE (6.5) : RECEIPTS, UTILISATION AND REPAYMENT PER HOUSEHOLD OF GOVERNMENT LOANS AND GRANTS FOR DIFFERENT PURPOSES IN ALL LOCALITIES

type of loans and grants	number of households=1349					
	number of house-holds	amount received (in Rs.)	extent utilised		amount repaid up to the period of survey (in Rs.)	interest paid up to the period of survey (in Rs.)
			amount (Rs).	per-centage		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. house-building	3	4750.0	4500.0	94.7	—	—
2. trade	350	1045.5	813.7	77.8	7.9	7.3
3. profession	29	664.7	504.0	75.8	0.2	—
4. maintenance	14	561.4	561.4	100.0	102.1	—
5. education loan	1	1500.0	1500.0	100.0	600.0	—
6. maintenance grant	8	455.5	262.5	57.6	—	—
7. education grant	2	133.0	133.0	100.0	—	—

6.14. The figures above show the position for all displaced households living in government colonies and in other localities taken together, and are thus derived from the figures in Tables (6.3) and (6.4). Since detailed comments have been given for the figures of the two earlier tables, further comments will not perhaps be necessary.

6.15. A distribution of the households arranged by the amounts of government loans (of all types) received by them is given in Table (6.6) below.

TABLE (6.6) : DISTRIBUTION OF HOUSEHOLDS BY THE AMOUNT OF GOVERNMENT LOANS RECEIVED BY THEM

amount of loans received (in Rs.)	government colonies		other localities		all localities	
	number of households	percent-age	number of households	percent-age	number of households	percent-age
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. 0	559	63.6	396	84.4	955	70.9
2. up to 499	109	12.3	18	3.9	127	9.3
3. 500 — 999	150	17.1	23	4.9	173	12.8
4. 1000 — 2999	52	5.9	20	4.3	72	5.4
5. 3000 — 4999	1	0.1	2	0.4	3	0.2
6. 5000 — 9999	8	0.9	8	1.7	16	1.2
7. 10000—19999	1	0.1	1	0.2	2	0.1
8. 20000 & above	—	—	1	0.2	1	0.1
9. total	880	100.0	469	100.0	1349	100.0

6.16. Taking the households of all localities together, 70.9 per cent did not receive government loans of any kind. 9.3 per cent received loans below Rs.500 and 22.1 per cent below Rs.1000. Only 1.6 per cent of the households received loans amounting to more than Rs.3000.

6.17. In government colonies, 63.6 per cent of the households did not receive government loans of any kind. 12.3 per cent of the households received loans below Rs.500 and 29.4 per cent below Rs.1000. Only 1.1 per cent of the households got loans amounting to Rs.3000 or over.

6.18. In other localities, 84.4 per cent of the households did not receive government loans of any kind. 3.9 per cent of the households received loans below Rs.500 and 8.8 per cent below Rs.1000. The proportion of households who got loans amounting to Rs.3000 or over was 2.5 per cent.

6.19. Some percentages indicating the overall proportions of utilisation of government loans have been given in Tables (6.3)—(6.5) above. It may now be interesting to see the distribution of the households arranged by the degree of utilisation of the loans in purposes for which these were taken and the relevant figures are given in Table (6.7) below.

TABLE (6.7) : DISTRIBUTION OF HOUSEHOLDS RECEIVING GOVERNMENT LOANS BY PERCENTAGE OF UTILISATION FOR PURPOSES FOR WHICH TAKEN

	government colonies			other localities			all localities		
percentage of utilisation	number of households	percentage	cumulative percentage	number of households	percentage	cumulative percentage	number of households	percentage	cumulative percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. up to 10	69	21.5	100.0	10	13.7	100.0	79	20.1	100.0
2. 10.1 — 30	14	4.4	78.5	1	1.4	86.3	15	3.8	79.9
3. 30.1 — 50	9	2.8	74.1	2	2.7	84.9	11	2.8	76.1
4. 50.1 — 70	46	14.3	71.3	4	5.5	82.2	50	12.7	73.3
5. 70.1 — 90	19	5.9	57.0	7	9.5	76.7	26	6.6	60.6
6. 90.1—100	164	51.1	51.1	49	67.2	67.2	213	54.0	54.0
7. total	321	100.0		73	100.0		394	100.0	
8. number of households	880			469			1349		

6.20. Taking the households of all localities together, it will be seen that 54.0 per cent of the households utilised almost the full amount of their loans in purposes for which these were taken. At the other extreme, 20.1 per cent of the households utilised only up to 10 per cent. In the government colonies, 51.1 per cent of the households utilised almost full amounts of their loans in purposes for which these loans were taken. 21.5 per cent of the households utilised only up to

10 per cent. In other localities, 67.2 per cent of the households utilised almost the whole amount in purposes for which the loans were taken. 13.7 per cent of the households utilised only up to 10 per cent.

6.21. A separate table showing the percentages of utilisation of trade and professional loans is given in Table (6.8) below.

TABLE (6.8) : PERCENTAGE DISTRIBUTION OF HOUSEHOLDS RECEIVING TRADE AND PROFESSIONAL LOANS BY PERCENTAGE UTILISATION FOR THE PURPOSE TAKEN

	percentage of utilisation	government colonies		other localities		all localities	
		trade loans	prof. loans	trade loans	prof. loans	trade loans	prof. loans
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	up to 10	21.4	32.0	16.7	—	20.6	27.6
2.	10.1 — 30	4.3	—	—	—	3.4	—
3.	30.1 — 50	2.9	4.0	1.5	25.0	2.5	6.9
4.	50.1 — 70	15.2	16.0	6.0	—	13.5	13.8
5.	70.1 — 90	6.3	4.0	9.1	25.0	6.9	6.8
6.	90.1—100	49.9	44.0	66.7	50.0	53.1	44.9
7.	total	100.0	100.0	100.0	100.0	100.0	100.0
8.	number of households which got loans	284	25	66	4	350	29
9.	number of households	880		469		1349	

6.22. Taking all the households of all localities together, 53.1 per cent of the households utilised nearly fully the trade loans in trading. 20.6 per cent utilised only up to 10 per cent. As compared with this, 44.9 per cent of the households taking professional loans utilised the money almost to the full extent in profession and 27.6 per cent utilised only up to 10 per cent.

6.23. In government colonies, 49.9 per cent of the households taking trade loans utilised the money nearly fully in trading while 21.4 per cent utilised only up to 10 per cent. Regarding professional loans, 44.0 per cent of the households utilised nearly the whole money in profession while 32.0 per cent of the households utilised only up to 10 per cent.

6.24. In other localities, 66.7 per cent of the households who took trade loans utilised the money nearly fully in trade; and 16.7 per cent of the households utilised only up to 10 per cent. Only 4 households got professional loans and half of them almost fully utilised the loans in setting up their professions.

CHAPTER SEVEN

HOUSING AND MISCELLANEOUS STATISTICS

HOUSING

7.1. Detailed figures regarding the housing conditions of the displaced households are given in Tables (25) to (31) at the end of this report, but the more important results have been discussed in this chapter. The Table (7.1) below shows the distribution of the sample households by types of possession of their living accommodation.

TABLE (7.1) : DISTRIBUTION OF HOUSEHOLDS BY TYPES OF POSSESSION OF LIVING ACCOMMODATION

nature of occupation	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. rented	869	98.9	409	89.3	1278	95.6
2. owned	10	1.1	49	10.7	59	4.4
3. total	879	100.0	458	100.0	1337	100.0
4. no information	1	—	11	—	12	—
5. sample households	880		469		1349	

7.2. Taking the households of all localities together and for which the information is available, 95.6 per cent were living in rented accommodation. In the government colonies, the proportion of households living in rented accommodation was 98.9 per cent; and for other localities the proportion was 89.3 per cent.

7.3. A distribution of households by the number of rooms occupied by them is given in Table (7.2) below.

TABLE (7.2) : DISTRIBUTION OF HOUSEHOLDS BY THE NUMBER OF ROOMS OCCUPIED

number of rooms	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. 1	753	85.7	192	41.0	946	70.2
2. 2	91	10.3	174	37.2	264	19.6
3. 3	23	2.6	66	14.1	89	6.6
4. 4 and above	12	1.4	36	7.7	48	3.6
5. total	879	100.0	468	100.0	1347	100.0
6. no information	1	—	1	—	2	—
7. sample households	880		469		1349	
8. average person per room		4.0		2.9		3.6

7.4. Taking all localities together and those households for whom the information is available, 70.2 per cent were living in one-room accommodation, 19.6 per cent in two-room accommodation and 10.2 per cent had three or more rooms to accommodate them. On average, 3.6 persons were living in a room.

7.5. As it will be seen, the households living outside the government colonies were in a better position from the point of view of housing also. 2.9 persons were living in a room as against 4.0 persons per room in the government colonies. 41.0 per cent were living in one-room accommodation as compared with 85.7 per cent in the government colonies; and this is not the complete picture. Many of the households shown as occupying one-room accommodation in government colonies had rooms in notional sense rather than in real sense because they were living in large sheds together with other households without any privacy. For instance, in Ulhasnagar township about a third of the households were living in big sheds where 20 to 30 households were accommodated under one roof with no proper partitions. In some cases partitions were put up by the households themselves with bamboo matting, wood or similar material; and there were cases where the partitioning was done by sackcloth, and in many cases there were no screens at all.

7.6. It is regretted that specific information regarding the proportion of households living under these conditions was not collected in this survey because it was not realised at the time of preparing the questionnaire that even after four or five years since migration, the housing conditions might be like this. As there were no clear instructions on this point, the investigators took as one room the floor-space occupied by a household in such a shed and even in recording the type of plinth, wall and roof, they recorded the particulars for the shed as a whole. It has been reported, however, that these sheds in the colonies are being reconditioned and partitioned into one-room tenements.

7.7. The distribution of households by the area of floor space occupied by them is given in Table (7.3) below.

TABLE (7.3) : DISTRIBUTION OF HOUSEHOLDS BY THE AREA OF FLOOR SPACE OCCUPIED

floor space (sq. ft.)	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. up to 100	70	7.9	80	17.0	150	11.1
2. 101—200	468	53.2	146	31.1	614	45.5
3. 201—300	218	24.8	112	23.9	330	24.5
4. 301—400	64	7.3	59	12.6	123	9.1
5. 401 & above	60	6.8	72	15.4	132	9.8
6. total	880	100.0	469	100.0	1349	100.0
7. floor space per person (sq. ft.)		46.0		47.9		46.7

7.8. Taking all localities together, only 11.1 per cent of the households had floor space of 100 sq. ft. or even less than that. 45.5 per cent of households had floor space between 101 and 200 sq. ft. and 24.5 per cent floor space between 201 and 300 sq. ft. 18.9 per cent had floor area over 300 sq. ft. The average floor space per person was 46.7 sq. ft.

7.9. In government colonies, the main concentration of the households was in areas between 101 and 300 sq. ft. accounting for 78.0 per cent. 7.9 per cent had floor space of 100 sq. ft. or below; and 14.1 per cent had floor areas above 300 sq. ft. The average floor space per person was 46.0 sq. ft.

7.10. In other localities, the distribution was more spread out. 17.0 per cent of the households had floor space of 100 sq. ft. or below, 31.1 per cent between 101 and 200 sq. ft. and 23.9 per cent between 201 and 300 sq. ft. 28.0 per cent of the households had floor space over 300 sq. ft. The average floor space per person was 47.9 sq. ft.

7.11. The distribution of the households by the nature of the structures in which they were living is given in Table (7.4) below.

TABLE (7.4) : DISTRIBUTION OF HOUSEHOLDS BY TYPE OF CONSTRUCTION OF LIVING ACCOMMODATION

type of construction	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. totally brick-built	81	9.2	137	29.2	218	16.2
2. mixed	710	80.7	156	33.3	866	64.2
3. totally non-brick-built	89	10.1	176	37.5	265	19.6
4. total	880	100.0	469	100.0	1349	100.0

7.12. Taking all localities together, 16.2 per cent of households were living in totally brick-built structures, 64.2 per cent in partly brick-built and partly non-brick-built structures and 19.6 per cent in totally non-brick-built ones.

7.13. In government colonies, 9.2 per cent of households were living in totally brick-built structures as compared with a proportion of 29.2 per cent for other localities. The houses in the government colonies were somewhat standardised in the sense that 80.7 per cent of the households were living in structures of mixed type. As compared with this, the proportion living in mixed structures in other localities was 33.3 per cent. 10.1 per cent of the households in the government colonies were living in totally non-brick-built structures as compared with 37.5 per cent for other localities.

7.14. A distribution of households by type of plinth, wall and roof of living accommodation is given in Table (7.5) below.

TABLE (7.5) : DISTRIBUTION OF HOUSEHOLDS BY TYPE OF CONSTRUCTION OF LIVING ACCOMMODATION

type of construction				government colonies		other localities		all localities	
				number of house-holds	percent-age	number of house-holds	percent-age	number of house-holds	percent-age
(1)				(2)	(3)	(4)	(5)	(6)	(7)
	plinth	*wall*	*roof*						
1.	brick	brick	brick	81	9.2	137	29.2	218	16.2
2.	brick	brick	tile, c.i. sheets, etc.	662	75.2	140	29.8	802	59.5
3.	brick	c.i. sheets	c.i. sheets	11	1.3	8	1.7	19	1.4
4.	mud	brick	tile, c.i. sheets, etc.	37	4.2	8	1.8	45	3.3
5.	other	other	other	89	10.1	176	37.5	265	19.6
6.	total			880	100.0	469	100.0	1349	100.0

7.15. Most of the structures in the government colonies had brick-built plinths, brick-built walls and roofs made of tiles, c.i. sheets or similar material. The households living in such structures account for 75.2 per cent of the total. As compared with this, the proportion of households living in such structures in other localities was 29.8 per cent. Next in order of importance were the totally non-brick-built structures; 10.1 per cent of the households in the government colonies and 37.5 per cent of the households in other localities were living in such structures.

LIGHTING

7.16. The distribution of households by the type of lights used by them is given in Table (7.6) below.

TABLE (7.6) : DISTRIBUTION OF HOUSEHOLDS BY TYPE OF LIGHTING

type of lighting	government colonies		other localities		all localities	
	number of house-holds	percent-age	number of house-holds	percent-age	number of house-holds	percent-age
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. electric	326	37.1	334	71.5	660	49.1
2. kerosene	542	61.7	133	28.5	675	50.1
3. others	11	1.2	—	—	11	0.8
4. total	879	100.0	467	100.0	1346	100.0
5. no information	1		2		3	
6. sample households	880		469		1349	

7.17. Taking all localities together and all the households for which the information is available, 49.1 per cent were using electric lights and 50.1 per cent kerosene lights. In government colonies 37.1 per cent of the households were using electric lights and 61.7 per cent lights with kerosene oil. A very small proportion of households, e.g., 1.2 per cent used other oils. In other localities, 71.5 per cent of the households were using electric lights and 28.5 per cent used kerosene oil for lights.

Water supply

7.18. In regard to drinking water the distribution of households by source of water and the average distance of the source from the living quarter is shown in Table (7.7) below.

TABLE (7.7) : DISTRIBUTION OF HOUSEHOLDS BY SOURCE OF DRINKING WATER WITH AVERAGE DISTANCE OF SOURCE

	government colonies			other localities			all localities		
source of drinking water	number of households	percentage	distance in yds.	number of households	percentage	distance in yds.	number of households	percentage	distance in yds.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. tube well	2	0.2	22.5	1	0.2	10.0	3	0.2	18.3
2. tap	729	82.9	22.4	394	84.1	11.5	1123	83.2	18.6
3. well	147	16.7	43.8	71	15.1	58.8	218	16.2	48.7
4. ers	2	0.2	—*	3	0.6	136.7	5	0.4	82.0
5. total	880	100.0	25.9	469	100.0	19.5	1349	100.0	23.7

* not available

7.19. Over 80 per cent of the households got their drinking water from taps. In government colonies the proportion of such households was 82.9 per cent and in other localities it was 84.1 per cent. 16.7 per cent of the households in government colonies got drinking water from wells as compared with 15.1 per cent in other localities. Tube wells as a source of supply of water were rare.

7.20. The average distance of the source was longer in government colonies than in other localities. The households living in government colonies got drinking water from taps which on average were located 22.4 yards away as compared with 11.5 yards for other localities. In the case of wells, the average distance in government colonies was 43.8 yards as compared with 58.8 yards for other localities.

7.21. The same source of water was used by a number of households and the data in this respect are given in Table (7.8) below.

TABLE (7.8) : DISTRIBUTION OF HOUSEHOLDS BY NUMBER USING THE SAME SOURCE OF DRINKING WATER

number of households using the same source	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. 1	63	7.2	171	40.2	234	18.0
2. 2	6	0.7	28	6.6	34	2.6
3. 3	2	0.2	19	4.5	21	1.6
4. 4	11	1.3	15	3.5	26	2.0
5. 5—10	36	4.1	74	17.4	110	8.5
6. 11—20	131	15.0	37	8.7	168	12.9
7. 21—30	192	22.0	19	4.5	211	16.3
8. above 30	432	49.5	62	14.6	494	38.1
9. total	873	100.0	425	100.0	1298	100.0
10. no information	7	—	44	—	51	—
11. sample households	880	—	469	—	1349	—

7.22. In government colonies, only 7.2 per cent of the households did not share the use of the source with other households. This proportion for other localities was 40.2 per cent. At the other extreme, 49.5 per cent of the households living in government colonies were sharing the use of the source with more than 30 households; the proportion for households living in other localities was 14.6 per cent. If the proportion of households sharing the source with more than 10 others is considered, the percentage for government colonies comes to 86.5 as compared with 27.8 per cent for other localities.

7.23. Information was also collected in regard to water used for purposes other than drinking. Table (7.9) below shows the distribution of the households by the source.

TABLE (7.9) : DISTRIBUTION OF HOUSEHOLDS BY SOURCE OF NON-DRINKING WATER WITH AVERAGE DISTANCE OF SOURCE

source of non-drinking water	government colonies			other localities			all localities		
	number of households	percentage	distance in yds.	number of households	percentage	distance in yds.	number of households	percentage	distance in yds.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. tube well	2	0.2	22.5	1	0.3	10.0	3	0.3	18.3
2. tap	655	82.3	19.5	291	80.4	9.6	946	81.7	16.4
3. tank	2	0.2	125.0	11	3.0	293.6	13	1.1	267.7
4. well	127	16.0	42.9	50	13.8	55.7	177	15.3	46.5
5. others	10	1.3	32.0	9	2.5	224.4	19	1.6	123.2
6. total	796	100.0	21.4	362	100.0	23.1	1158	100.0	22.0
7. no information	84			107			191		
8. sample households	880			469			1349		

7.24. It will be seen that information from as many as 191 households is not available. When these households are excluded from the sample and the rest are considered, the distribution of the households comes out broadly the same as that in the case of drinking water, except that a few households supplemented their water supply from tanks. The average distances of the sources are also similar to those recorded for drinking water.

7.25. Information regarding the number of households using the same source for non-drinking water is given in Table (7.10) below.

TABLE (7.10) : DISTRIBUTION OF HOUSEHOLDS BY NUMBERS USING THE SAME SOURCE FOR NON-DRINKING WATER

number of households using the same source	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. 1	56	7.9	125	39.6	181	17.7
2. 2	6	0.8	21	6.7	27	2.6
3. 3	1	0.1	17	5.4	18	1.8
4. 4	10	1.4	15	4.8	25	2.4
5. 5—10	23	3.3	50	15.9	73	7.1
6. 11—20	94	13.3	28	8.9	122	11.9
7. 21—30	166	23.4	17	5.4	183	17.9
8. above 30	353	49.8	42	13.3	395	38.6
9. total	709	100.0	315	100.0	1024	100.0
10. no information	171	—	154	—	325	—
11. sample households	880	—	469	—	1349	—

7.26. The pattern of crowding in sharing a common source is more or less similar to that observed in the case of drinking water.

LATRINES

7.27. The distribution of households by types of latrines used by them together with the average distance of the latrines from the living quarters is given in Table (7.11) below.

TABLE (7.11) : DISTRIBUTION OF HOUSEHOLDS BY TYPE OF LATRINES USED WITH AVERAGE DISTANCE OF THE LATRINES

type of latrines	government colonies			other localities			all localities		
	number of households	percentage	distance in yds.	number of households	percentage	distance in yds.	number of households	percentage	distance in yds.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. underground drainage with flush	27	3.1	3.9	219	46.9	6.1	246	18.3	5.9
2. septic tank	149	17.0	26.9	6	1.3	29.2	155	11.6	27.0
3. old type	685	78.4	31.0	198	42.4	24.6	883	65.8	29.6
4. others	13	1.5	93.8	44	9.4	146.1	57	4.3	134.2
5. total	874	100.0	30.2	467	100.0	27.3	1341	100.0	29.2
6. no information	6	—	—	2	—	—	8	—	—
7. sample households	880	—	—	469	—	—	1349	—	—

7.28. In government colonies, 3.1 per cent of the households were using latrines with underground drain and flush. 17.0 per cent were using latrines of septic tank type, 78.4 per cent of the old type latrines with pots or baskets. Latrines used by the remaining households were unspecified. In other localities, 46.9 per cent were using latrines with underground drains and flush. 1.3 per cent were using latrines of septic tank type and 42.4 per cent the old type latrines. 9.4 per cent of the households were using latrines of other types or just fields.

7.29. The average distance, taking all types of latrines together, was 30.2 yards in government colonies and 27.3 yards in other localities. The average distance varied between types of latrines. Latrines with underground drains and flush were nearer. Septic tank and old type latrines were located at a distance averaging between 27 to 29 yards.

7.30. Information as to how many households were using the same latrines is given below in Table (7.12). The figures relate to structures. The number of seats in a structure varied usually between 2 to 12 seats.

TABLE (7.12) : DISTRIBUTION OF HOUSEHOLDS BY NUMBERS USING THE SAME LATRINES (STRUCTURES)

number of households using the same structure	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. 1	38	4.4	116	27.7	154	12.0
2. 2	32	3.7	51	12.1	83	6.5
3. 3	5	0.6	32	7.6	37	2.9
4. 4	90	10.4	30	7.1	120	9.3
5. 5—10	35	4.1	98	23.3	133	10.4
6. 11—20	93	10.8	49	11.7	142	11.1
7. 21—30	145	16.8	20	4.8	165	12.9
8. above 30	426	49.2	24	5.7	450	34.9
9. total	864	100.0	420	100.0	1284	100.0
10. no information	16	—	49	—	65	—
11. sample households	880	—	469	—	1349	—

7.31. In government colonies, only 4.4 per cent of the households were using latrines without sharing them with others. In comparison, the proportion for other localities was 27.7 per cent. At the other end, 49.2 per cent in government colonies were using latrines shared by more than 30 households whereas this percentage for other localities was 5.7. In government colonies, 76.8 per cent of the households were using latrines sharing with more than 10 others, and the corresponding proportion for other localities was 22.2 per cent.

PREFERENCE REGARDING MEANS OF LIVELIHOOD

7.32. Some information on preference of the households regarding their means of livelihood was collected in the survey. Preferences of the households do not necessarily mean that they will be more successful in their preferred lines than their present lines of activities; but these particulars, probably, indicate to what extent they had been settling down to their current lines of activities. The information, however, related to the means of livelihood of the principal earners of the households only. A two-way distribution of households by present means of livelihood and those preferred is given in Table (39) at the end. A summary of the figures is given in Table (7.13) below.

TABLE (7.13) : NUMBER AND PROPORTION OF HOUSEHOLDS PREFERRING TO CONTINUE THE SAME MEANS OF LIVELIHOOD AS FOLLOWED AT THE TIME OF THE SURVEY

livelihood class	government colonies			other localities			all localities		
	present livelihood	preference for same		present livelihood	preference for same		present livelihood	preference for same	
	number of households	number of households	percentage	number of households	number of households	percentage	number of households	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. agriculture	2	1	50.0	1	1	100.0	3	2	66.7
2. crafts and cottage industries	84	44	52.4	40	28	70.0	124	72	58.1
3. large industries	30	15	50.0	24	16	66.7	54	31	57.4
4. trade	303	199	65.7	198	153	77.3	501	352	70.3
5. construction	18	10	55.6	9	4	44.4	27	14	51.9
6. transport	35	17	48.6	24	16	66.7	59	33	55.9
7. services	182	92	50.5	86	56	65.1	268	148	55.2
8. prof. services	49	29	59.2	27	16	59.3	76	45	59.2
9. others	155	95	61.3	56	27	48.2	211	122	57.8
10. total	858	502	58.5	465	317	68.2	1323	819	61.9
11. no information	22	—	—	4	—	—	26	—	—
12. sample households	880	—	—	469	—	—	1349	—	—

7.33. Taking the households of all localities together and all these for which information is available, 61.9 per cent preferred activities which they were following at the time of the survey. The proportions of households preferring to remain in the same activities varied between livelihood classes; the highest being in trade, namely, 70.3 per cent and the lowest in construction, namely, 51.9 per cent.

7.34. In government colonies, 58.5 per cent of the households preferred to remain in the same activities. The proportions for the different livelihood classes varied between 65.7 per cent for trade and 48.6 per cent for transport. In other localities, the proportion preferring to remain in the same activities was even higher, namely, 68.2 per cent, and over the different livelihood classes the proportions varied between 77.3 per cent for trade and 44.4 per cent for construction.

STATE PREFERENCE OF DISPLACED PERSONS

7.35. In regard to place of habitation the preferences of the households have been shown by the different States of India in Table (40) at the end and also in Table (7.14) below.

TABLE (7.14) : DISTRIBUTION OF HOUSEHOLDS BY STATES PREFERRED TO LIVE IN

State preferred	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. Uttar Pradesh	5	0.6	2	0.4	7	0.5
2. Bombay	581	66.1	389	83.0	970	71.9
3. Saurashtra	—	—	1	0.2	1	0.1
4. Madhya Pradesh	21	2.4	—	—	21	1.6
5. Madhya Bharat	3	0.3	—	—	3	0.2
6. Hyderabad	6	0.7	1	0.2	7	0.5
7. Bhopal	1	0.1	—	—	1	0.1
8. Rajasthan	3	0.3	—	—	3	0.2
9. Punjab	2	0.2	2	0.4	4	0.3
10. Delhi	8	0.9	7	1.5	15	1.1
11. no opinion	250	28.4	55	11.7	305	22.6
12. not recorded	—	—	12	2.6	12	0.9
13. total	880	100.0	469	100.0	1349	100.0

7.36. Taking the households of all localities together, 71.9 per cent stated that they preferred to live in the Bombay State. 22.6 per cent had no opinion. In other words, only 5.5 per cent of the households preferred to go out of Bombay. In government colonies, 66.1 per cent of the households preferred to remain in the Bombay State and 28.4 per cent had no opinion. That is, only 5.5 per cent of the households preferred to go to other States, and of these nearly half preferred Madhya Pradesh.

7.37. In other localities, 83.0 per cent of the households preferred to stay in Bombay and 11.7 per cent had no opinion. Another 2.6 per cent of the households preferred to go away from Bombay. Many of the households living in government colonies as well as in other localities preferred change of places within the Bombay State and the general preference was larger towns than where they had been living.

Knowledge of local language

7.38. Information was also collected in regard to the extent to which the local languages were known to the households. The sample households have been arranged in Table (7.15) by five categories, namely, (1) those who can read, write and speak the local language, (2) those who can read and speak but cannot write the language, (3) those who can speak and cannot read or write the language, (4) those who can understand but cannot read, write or speak the local language, and, finally, (5) those who cannot read, write, speak or understand the local language. It should be mentioned in this connection that the question was limited to local languages only, mainly Marathi or Gujrati, depending on the area in which the households were living. The majority of the households, however, had working knowledge of Hindustani.

TABLE (7.15) : DISTRIBUTION OF HOUSEHOLD HEADS BY DEGREE OF KNOWLEDGE OF LOCAL LANGUAGE

knowledge of local language	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. can read, write and speak	46	5.3	83	17.8	129	9.7
2. can read and speak but cannot write	5	0.6	22	4.7	27	2.0
3. can speak but cannot read or write	99	11.4	80	17.1	179	13.4
4. understand but cannot read, write or speak	63	7.3	58	12.4	121	9.1
5. cannot read, write, speak or understand	655	75.4	224	48.0	879	65.8
6. total	868	100.0	467	100.0	1335	100.0
7. not returned	12	—	2	—	14	—
8. sample households	880	—	469	—	1349	—

7.39. Taking the households of all localities together, 65.8 per cent could not read, write, speak or understand the local language. The proportion for government colonies was 75.4 per cent. The proportion for other localities was much better, namely, 48.0 per cent. It may probably be repeated that the majority of these households, however, had working knowledge of Hindustani.

NATIONAL SAMPLE SURVEY : SURVEY OF DISPLACED PERSONS IN BOMBAY

LIST OF TABLES

PAGE

TABLE 1 : Distribution of sample households of displaced persons by division, district and census tract in Bombay State : July—September 1953 285

DEMOGRAPHY

TABLE 2 : Distribution of displaced households by mother tongue : July-September 1953 286

TABLE 3 : Distribution of members of sample households by age and sex : July-September 1953 287

TABLE 4 : Distribution of members of sample households by age, sex and education standard : July—September 1953 288

TABLE 5 : Average number of earners, earning dependents and non-earning dependents per household by sex : July—September 1953 290

TABLE 6 : Percentage of earners, earning dependents and non-earning dependents by sex : July—September 1953 290

TABLE 7 : Distribution of economically active persons by livelihood classes and employment status : July—September 1953 291

TABLE 8 : Distribution of economically active persons by livelihood classes with average number of days employed and average income earned per month : July—September 1953 292

TABLE 9 : Distribution of economically active persons by days of employment in a month with average income earned : July—September 1953 292

TABLE 10 : Distribution of sample households by livelihood classes before migration and in the period of the survey : July—September 1953 293

EXPENDITURE

TABLE 11 : Consumer expenditure per household per month by levels of total monthly expenditure : July—September 1953 294

INCOME

TABLE 12 : Average monthly receipts per household (over all households) distributed by source of receipts and monthly expenditure levels : July—September 1953 295

TABLE 13 : Number of households (effective samples) and average monthly receipts by sources of receipts and expenditure levels : July—September 1953 297

TABLE 14 : Distribution of sample households by level of total monthly receipts : July—September 1953 300

TABLE 15 : Distribution of households by percentage of income (from occupation, "home supply" and "other sources") to total expenditure : July—September 1953 300

TABLE 16 : Number of households using non-government loans for consumption in the period under investigation arranged by types of securities given with or without interest : July—September 1953 301

LOANS

TABLE 17 : Number of households having loans and amount of outstanding loans per household : July—September 1953 302

TABLE 18 : Distribution of households by levels of outstanding (government and non-government) loans : July—September 1953 302

TABLE 19 : Number of non-government loans obtained since migration with or without interest arranged by types of securities given : July—September 1953 303

PAGE

TABLE 20 : Number of non-government loans obtained since migration with or without interest by purposes for which these were taken and by types of securities given : July—September 1953 304

TABLE 21 : Amount of government loans and grants received per household and the extent of their utilisation ; amount of interest paid and loans repaid per household : July—September 1953 307

TABLE 22 : Distribution of households by total amounts of government loans : July—September 1953 309

TABLE 23 : Distribution of households by types and amounts of government loans : July—September 1953 310

TABLE 24 : Distribution of households by levels of utilisation in purposes for which the loans were granted : July—September 1953 312

HOUSING AND MISCELLANEOUS STATISTICS

TABLE 25 : Distribution of households by the nature of use of land under occupation : July—September 1953 313

TABLE 26 : Distribution of households by number of rooms and type of possession of living accommodation : July—September 1953 314

TABLE 27 : Percentage distribution of households by floor space and type of construction : July—September 1953 315

TABLE 28 : Percentage distribution of households by floor space and type of construction : July—September 1953 316

TABLE 29 : Distribution of households by floor space and size of households with percentage distribution by floor space : July—September 1953 317

TABLE 30 : Distribution of households by floor space and size of households with percentage distribution by size of households : July—September 1953 318

TABLE 31 : Distribution of households by type of construction of plinth, wall and roof of living accommodation : July—September 1953 319

TABLE 32 : Distribution of households by type of lighting : July—September 1953 .. 319

TABLE 33 : Distribution of households by source of drinking water with average distance of source : July—September 1953 320

TABLE 34 : Distribution of households by source of non-drinking water with average distance of source : July—September 1953 320

TABLE 35 : Distribution of households by type of latrines used with average distance of latrines : July—September 1953 320

TABLE 36 : Distribution of households by number using same source of drinking water : July—September 1953 321

TABLE 37 : Distribution of households by number using same source of non-drinking water : July—September 1953 322

TABLE 38 : Distribution of households by number using same latrine : July–September 1953 .. 323

TABLE 39 : Distribution of households by present livelihood classes : July–September 1953.. 324

TABLE 40 : Distribution of households by the States preferred for living in : July—September 1953 326

TABLE 41 : Distribution of heads of households by the extent of knowledge of local language : July—September 1953 326

TABLE 1 : DISTRIBUTION OF SAMPLE HOUSEHOLDS OF DISPLACED PERSONS BY DIVISION, DISTRICT AND CENSUS TRACT IN BOMBAY STATE : JULY—SEPTEMBER 1953

division	district	census tract	number of sample households		
			government colonies	other localities	all localities
(1)	(2)	(3)	(4)	(5)	(6)
1. Bombay Deccan, Northern	1. East Khandesh	1. Jalgaon, Bhusawal, Yawl and Ravat	27	–	27
		2. Amalner, Parola, Chopda and Erandol	4	–	4
		3. Chalisgaon, Jamner, Pachora and Bhadgaon	5	–	5
	2. Poona	1. Poona City Corporation	9	45	54
		3. Baramati, Indapur, Dhond and Purandhar	1	5	6
	3. Nasik	1. Nasik and Igatpuri	4	27	31
	4. Ahmednagar	1. Ahmednagar, Saigonda, Karjet and Jamkhed	10	16	26
2. Bombay, Gujrat	1. Mehsana	1. Bijapur, Visnagar, Kaloi and Kadi	1	3	4
	3. Panch Mahals	1. Lonawada, Kaloi, Haloi, Gadhra and Santrampur	1	15	16
	4. Ahmedabad	1. Ahmedabad City	21	63	84
		2. Dholka, Dhandhuka, Ahmedabad City Taluqas and Sonandapeta	117	18	135
	5. Surat	1. Surat City	–	5	5
	6. Banaskantha	1. Tharad, Deesu, Palanpur and Abu Road	–	8	8
	7. Baroda	1. Baroda City	13	42	55
3. Greater Bombay	1. Greater Bombay	1. Colaba, Fort, Boribunder, Mandvi, Dongri and Umarkhadi wards	–	16	16
		2. Khara Talao, Kumbharwada Market, Dhobi Talao and Fanaswadi wards	1	7	8
		3. Girgaon, Khetwadi, Walkeswar, Mahalakshmi and Tardeo wards	–	9	9
		4. Kamathipura, Nagpada, Mazagaon, Byculla West and Byculla East wards	–	3	3
		5. Gherapedso, Sewri, Lalbag, Parel and Naigaum wards	–	13	13
		6. Matunga, Sion, Lovegroove, Lower Parel and Worli wards	12	64	76
		7. Parbhadevi, Dadar and Mahim wards	1	34	35
		8. Bandra, Parli, Andheri and Juhu wards	4	21	25
		9. Ghatkopar, Kurla wards and 38 villages including Chimur	35	37	72
4. Bombay, Konkan	2. Thana	1. Thana, Bonvali and Kalyan	614	18	632
5. total			880	469	1349

TABLE 2 : DISTRIBUTION OF DISPLACED HOUSEHOLDS BY MOTHER TONGUE : JULY—SEPTEMBER 1953

mother tongue	government colonies		other localities		all localities	
	number of house-holds	percent-age	number of house-holds	percent-age	number of house-holds	percent-age
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. marathi	13	1.5	6	1.3	19	1.4
2. sindhi	777	88.3	315	67.2	1092	81.0
3. punjabi	24	2.7	76	16.2	100	7.4
4. urdu	1	0.1	2	0.4	3	0.2
5. hindi	6	0.7	4	0.9	10	0.8
6. gujrati	35	4.0	49	10.4	84	6.2
7. others	15	1.7	15	3.2	30	2.2
8. not returned	9	1.0	2	0.4	11	0.8
9. total	880	100.0	469	100.0	1349	100.0

TABLE 3 : DISTRIBUTION OF MEMBERS OF SAMPLE HOUSEHOLDS BY AGE AND SEX : JULY—SEPTEMBER 1953

age group	number of persons			percentage		
	male	female	total	male	female	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			government colonies			
0	38	47	85	1.8	2.2	2.0
1 – 4	211	230	441	9.8	10.8	10.3
5–14	609	558	1167	28.3	26.0	27.1
15–24	445	446	891	20.7	20.9	20.8
25–34	306	301	607	14.2	14.1	14.2
35–44	234	211	445	10.9	9.9	10.4
45–54	139	155	294	6.5	7.2	6.9
55–64	106	122	228	4.9	5.7	5.3
65–74	53	64	117	2.5	3.0	2.7
75 & above	9	4	13	0.4	0.2	0.3
total	2150	2138	4288	100.0	100.0	100.0
			other localities			
0	25	25	50	1.9	1.9	1.9
1 – 4	138	117	255	10.3	9.1	9.7
5–14	388	344	732	28.9	26.7	27.9
15–24	247	278	525	18.5	21.6	20.0
25–34	191	188	379	14.3	14.6	14.4
35–44	148	143	291	11.1	11.1	11.1
45–54	112	88	200	8.4	6.8	7.6
55–64	49	53	102	3.7	4.1	3.9
65–74	36	46	82	2.7	3.6	3.1
75 & above	3	7	10	0.2	0.5	0.4
total	1337	1289	2626	100.0	100.0	100.0
			all localities			
0	63	72	135	1.8	2.1	2.0
1 – 4	349	347	696	10.0	10.1	10.1
5–14	997	902	1899	28.6	26.2	27.4
15–24	692	724	1416	19.8	21.0	20.5
25–34	497	489	986	14.2	14.3	14.3
35–44	382	354	736	11.0	10.3	10.6
45–54	251	243	494	7.2	7.4	7.1
55–64	155	175	330	4.5	5.1	4.8
65–74	89	110	199	2.6	3.2	2.9
75 & above	12	11	23	0.3	0.3	0.3
not returned	–	–	–	–	–	–
total	3487	3427	6914	100.0	100.0	100.0

TABLE 4 : DISTRIBUTION OF MEMBERS OF SAMPLE HOUSEHOLDS BY AGE, SEX AND EDUCATION STANDARD : JULY—SEPTEMBER 1953

age group	education standard																				
	illiterate			merely literate			under matric			matric			graduate			not specified			total		
	male	female	total	male	female	total	male	female	total	male	female	total	male	female	total	male	female	total	male	female	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
										government colonies											
0	38	47	85	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	38	47	85
1–4	202	229	431	9	1	10	–	–	–	–	–	–	–	–	–	–	–	–	211	230	441
5–14	235	256	491	321	262	583	53	40	93	–	–	–	–	–	–	–	–	–	609	558	1167
15–24	70	197	267	184	182	366	127	54	181	59	12	71	5	1	6	–	–	–	445	446	891
25–34	49	177	226	161	104	265	55	18	73	37	2	39	4	–	4	–	–	–	306	301	607
35–44	44	155	199	125	49	174	42	7	49	18		18	4	–	4	1	–	1	234	211	445
45–54	38	118	156	77	35	112	14	2	16	4	–	4	6	–	6	–	–	–	139	155	294
55–64	35	104	139	54	17	71	9	1	10	6	–	6	2	–	2	–	–	–	106	122	228
65–74	21	58	79	27	6	33	4	–	4	1	–	1	–	–	–	–	–	–	53	64	117
75 & above	3	4	7	5	–	5	–	–	–	1	–	1	–	–	–	–	–	–	9	4	13
total	735	1345	2080	963	656	1619	304	122	426	126	14	140	21	1	22	1	–	1	2150	2138	4288
p.c.	34.2	62.9	48.5	44.8	30.7	37.8	14.1	5.7	9.9	5.9	0.7	3.3	1.0	0.0	0.5	0.0	–	0.0	100.0	100.0	100.0

TABLE 4 (*Contd.*) : DISTRIBUTION OF MEMBERS OF SAMPLE HOUSEHOLDS BY AGE, SEX AND EDUCATION STANDARD : JULY—SEPTEMBER 1953

age group	education standard																				
	illiterate			merely literate			under matric			matric			graduate			not specified			total		
	male	female	total	male	female	total	male	female	total	male	female	total	male	female	total	male	female	total	male	female	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
other localities																					
0	25	25	50	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	25	25	50
1 – 4	137	111	248	1	6	7	–	–	–	–	–	–	–	–	–	–	–	–	138	117	255
5–14	110	109	219	214	181	395	63	54	117	1	–	1	–	–	–	–	–	–	388	344	732
15–24	14	56	70	84	124	208	108	67	175	39	27	66	2	4	6	–	–	–	247	278	525
25–34	11	63	74	69	84	153	50	23	73	47	10	57	14	8	22	–	–	–	191	188	379
35–44	15	61	76	56	66	122	38	7	45	27	8	35	12	–	12	–	1	1	148	143	291
45–54	12	53	65	49	29	78	27	5	32	18	1	19	6	–	6	–	–	–	112	88	200
55–64	9	39	48	25	13	38	7	1	8	5	–	5	3	–	3	–	–	–	49	53	102
65–74	9	35	44	18	9	27	5	1	6	3	1	4	1	–	1	–	–	–	36	46	82
75 & above	1	6	7	2	1	3	–	–	–	–	–	–	–	–	–	–	–	–	3	7	10
total	343	558	901	518	513	1031	298	158	456	140	47	187	38	12	50	–	1	1	1337	1289	2626
p.c.	25.7	43.3	34.3	38.7	39.8	39.3	22.3	12.3	17.4	10.5	3.6	7.1	2.8	0.9	1.9	–	0.0	0.0	100.0	100.0	100.0
all localities																					
0	63	72	135	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	63	72	135
1 – 4	339	340	679	10	7	17	–	–	–	–	–	–	–	–	–	–	–	–	349	347	696
5–14	345	365	710	535	443	978	116	94	210	1	–	1	–	–	–	–	–	–	997	902	1899
15–24	84	253	337	268	306	574	235	121	356	98	39	137	7	5	12	–	–	–	692	724	1416
25–34	60	240	300	230	188	418	105	41	146	84	12	96	18	8	26	–	–	–	497	489	986
35–44	59	216	275	181	115	296	80	14	94	45	8	53	16	–	16	1	1	2	382	354	736
45–54	50	171	221	126	64	190	41	7	48	22	1	23	12	–	12	–	–	–	251	243	494
55–64	44	143	187	79	30	109	16	2	18	11	–	11	5	–	5	–	–	–	155	175	330
65–74	30	93	123	45	15	60	9	1	10	4	1	5	1	–	1	–	–	–	89	110	199
75 & above	4	10	14	7	1	8	–	–	–	1	–	1	–	–	–	–	–	–	12	11	23
total	1078	1903	2981	1481	1169	2650	602	280	882	266	61	327	59	13	72	1	1	2	3487	3427	6914
p.c.	30.9	55.5	43.1	42.5	34.1	38.4	17.3	8.2	12.8	7.6	1.8	4.7	1.7	0.4	1.0	0.0	0.0	0.0	100.0	100.0	100.0

TABLE 5 : AVERAGE NUMBER OF EARNERS, EARNING DEPENDENTS AND NON-EARNING DEPENDENTS PER HOUSEHOLD BY SEX : JULY—SEPTEMBER 1953

sex	government colonies				other localities				all localities			
	earner	earning dependent	dependent	total	earner	earning dependent	dependent	total	earner	earning dependent	dependent	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. male	0.87	0.12	1.45	2.44	1.05	0.17	1.63	2.85	0.93	0.14	1.51	2.58
2. female	0.05	0.04	2.34	2.43	0.04	0.03	2.67	2.74	0.05	0.03	2.46	2.54
3. total	0.92	0.16	3.79	4.87	1.09	0.20	4.30	5.59	0.98	0.17	3.97	5.12
4. number of households	880				469				1349			

TABLE 6 : PERCENTAGE OF EARNERS, EARNING DEPENDENTS AND NON-EARNING DEPENDENTS BY SEX : JULY—SEPTEMBER 1953

sex	government colonies				other localities				all localities			
	earner	earning dependent	dependent	total	earner	earning dependent	dependent	total	earner	earning dependent	dependent	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. male	17.8	2.6	29.8	50.2	18.7	3.0	29.2	50.9	18.1	2.7	29.6	50.4
2. female	1.1	0.7	48.0	49.8	0.8	0.5	47.8	49.1	1.0	0.7	47.9	49.6
3. total	18.9	3.3	77.8	100.0	19.5	3.5	77.0	100.0	19.1	3.4	77.5	100.0
4. number of households	4288				2626				6914			

TABLE 7: DISTRIBUTION OF ECONOMICALLY ACTIVE PERSONS BY LIVELIHOOD CLASSES AND EMPLOYMENT STATUS: JULY—SEPTEMBER 1953

	employment status																	
	government colonies						other localities						all localities					
livelihood classes	employer	employee	own account worker	unearned income receiver	total	percentage to total	employer	employee	own account worker	unearned income receiver	total	percentage to total	employer	employee	own account worker	unearned income receiver	total	percentage to total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
1. agriculture, animal husbandry etc.	–	–	3	–	3	0.3	–	–	–	1	1	0.2	–	–	3	1	4	0.3
2. crafts and cottage industries	–	32	85	3	120	12.6	3	14	44	–	61	10.1	3	46	129	3	181	11.6
3. large scale industries	1	33	6	–	40	4.2	1	26	4	–	1	5.1	2	59	10	–	71	4.6
4. trade	4	56	339	–	399	42.1	9	56	221	1	287	47.4	13	112	560	1	686	44.1
5. construction	–	18	9	–	27	2.8	–	12	1	–	13	2.1	–	30	10	–	40	2.6
6. transport	–	44	3	–	47	4.9	–	30	2	–	32	5.3	–	74	5	–	79	5.1
7. services	2	172	70	1	245	25.8	4	93	37	1	135	22.3	6	265	107	2	380	24.4
8. professional services	–	25	34	–	59	6.2	–	16	20	–	36	6.0	–	41	54	–	95	6.1
9. others	–	–	–	10	10	1.1	–	1	3	5	9	1.5	–	1	3	15	19	1.2
10. total	7	380	549	14	950	100.0	17	248	332	8	605	100.0	24	628	881	22	1555	100.0
11. percentage	0.7	40.0	57.8	1.5	100.0		2.8	41.0	54.9	1.3	100.0		1.5	40.4	56.7	1.4	100.0	

TABLE 8: DISTRIBUTION OF ECONOMICALLY ACTIVE PERSONS BY LIVELIHOOD CLASSES WITH AVERAGE NUMBER OF DAYS EMPLOYED AND AVERAGE INCOME EARNED PER MONTH: JULY—SEPTEMBER 1953

livelihood classes	government colonies			other localities			all localities		
	number of persons	average of days employed	income per person in Rs.	number of persons	average of days employed	income per person in Rs.	number of persons	average of days employed	income per person in Rs.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. cultivator, agricultural and allied subjects	3	24	36.50	1	—	300.00	4	18	102.38
2. crafts and cottage industries	120	23	47.59	61	24	86.88	181	23	60.83
3. large-scale industries	40	24	100.56	31	23	141.60	71	24	118.48
4. trade	399	25	63.43	287	25	116.00	686	25	85.42
5. construction	27	20	96.77	13	22	187.27	40	21	126.18
6. transport	47	25	107.13	32	25	132.03	79	25	117.22
7. services	245	25	76.58	135	24	108.25	380	25	87.83
8. prof. services	59	22	95.54	36	23	157.29	95	22	118.82
9. others	10	—	100.22	9	3	139.55	19	2	118.85
10. total	950	24	71.78	605	24	118.14	1555	24	89.81

TABLE 9: DISTRIBUTION OF ECONOMICALLY ACTIVE PERSONS BY DAYS OF EMPLOYMENT IN A MONTH WITH AVERAGE INCOME EARNED: JULY—SEPTEMBER 1953

days of employment in the month	government colonies			other localities			all localities		
	number of persons	percentage	income per person in Rs.	number of persons	percentage	income per person in Rs.	number of persons	percentage	income per person in. Rs.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. 1–7	11	1.2	14.18	8	1.3	24.25	19	1.2	18.42
2. 8–14	28	2.9	30.67	6	1.0	46.08	34	2.2	33.39
3. 15–21	120	12.6	42.08	37	6.1	77.14	157	10.1	50.34
4. 22–30	791	83.3	78.54	554	91.6	123.01	1345	86.5	96.86
5. total	950	100.0	71.78	605	100.0	118.14	1555	100.0	89.81

NATIONAL SAMPLE SURVEY : SURVEY OF DISPLACED PERSONS IN BOMBAY

TABLE 10 : DISTRIBUTION OF SAMPLE HOUSEHOLDS BY LIVELIHOOD CLASSES BEFORE MIGRATION AND IN THE PERIOD OF THE SURVEY : JULY—SEPTEMBER 1953

present livelihood	livelihood before migration										
	agriculture, animal husbandry etc.	crafts and cottage industry	large-scale industries	trade	construction	transport	services	professional services	others	not returned	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					government colonies						
1. agriculture, animal husbandry etc.	1	–	–	1	–	–	–	–	–	–	2
2. crafts and cottage industries	13	41	2	13	2	–	9	1	2	1	84
3. large-scale industries	6	2	8	5	–	1	5	–	3	–	30
4. trade	49	4	5	205	2	2	28	3	5	–	303
5. construction	–	1	–	2	13	–	2	–	–	–	18
6. transport	4	2	–	10	–	9	6	2	2	–	35
7. services	22	12	4	40	4	3	83	3	10	1	182
8. professional services	4	–	–	2	–	–	2	38	3	–	49
9. others	35	10	4	50	3	1	24	8	11	9	155
10. not returned	5	4	1	5	–	1	3	1	1	1	22
11. total	139	76	24	333	24	17	162	56	37	12	880
					other localities						
1. agriculture, animal husbandry etc.	–	–	–	1	–	–	–	–	–	–	1
2. crafts and cottage industries	–	28	–	8	–	–	4	–	–	–	40
3. large-scale industries	1	1	7	11	–	–	2	–	2	–	24
4. trade	13	3	5	147	1	1	20	3	4	1	198
5. construction	–	–	–	1	6	–	1	1	–	–	9
6. transport	2	2	1	8	2	5	3	–	1	–	24
7. services	2	4	1	17	2	3	45	3	6	3	86
8. professional services	–	–	1	2	–	–	2	21	–	1	27
9. others	10	2	1	20	–	2	9	4	7	1	56
10. not returned	–	–	–	–	2	–	2	–	–	–	4
11. total	28	40	16	215	13	11	88	32	20	6	469
					all localities						
1. agriculture, animal husbandry etc.	1	–	–	2	–	–	–	–	–	–	3
2. crafts and cottage industries	13	69	2	21	2	–	13	1	2	1	124
3. large-scale industries	7	3	15	16	–	1	7	–	5	–	54
4. trade	62	7	10	352	3	3	48	6	9	1	501
5. construction	–	1	–	3	19	–	3	1	–	–	27
6. transport	6	4	1	18	2	14	9	2	3	–	59
7. services	24	16	5	57	6	6	128	6	16	4	268
8. professional services	4	–	1	4	–	–	4	59	3	1	76
9. others	45	12	5	70	3	3	33	12	18	10	211
10. not returned	5	4	1	5	2	1	5	1	1	1	26
11. total	167	116	40	548	37	28	250	88	57	18	1349

TABLE 11: CONSUMER EXPENDITURE PER HOUSEHOLD PER MONTH BY LEVELS OF TOTAL MONTHLY EXPENDITURE: JULY—SEPTEMBER 1953

area	items of expenditure	level of monthly expenditure											
		Rs. 1—50		Rs. 51–100		Rs. 101–200		Rs. 201–500		Rs. 501 & above		total	
		expenditure in Rs.	percentage	expenditure in Rs.	percentage	expenditure in Rs.	percentage	expenditure in Rs.	percentage	expenditure in Rs.	percentage	expenditure in Rs.	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	number of households	154		357		278		83		8		880	
	average size of households	2.28		4.48		6.01		7.22		9.38		4.88	
1. government colonies	1. food	25.77	72.8	53.59	70.3	88.96	65.0	153.99	56.0	279.91	43.2	71.42	63.7
	2. clothing	1.53	4.3	4.37	5.7	11.15	8.1	32.23	11.7	118.62	18.3	9.68	8.6
	3. house rent or maintenance	0.21	0.6	1.09	1.4	5.02	3.7	13.32	4.8	9.85	1.5	3.41	3.0
	4. fuel and light	3.33	9.4	6.08	8.0	8.15	5.9	13.23	4.8	20.75	3.2	7.06	6.3
	5. miscellaneous	4.57	12.9	11.13	14.6	23.70	17.3	62.59	22.7	219.00	33.8	20.70	18.4
	6. total	35.41	100.0	76.26	100.0	136.98	100.0	275.36	100.0	648.13	100.0	112.27	100.0
	number of households	20		76		213		137		23		469	
	average size of households	2.25		3.70		5.51		7.04		6.83		5.59	
2. other localities	1. food	24.21	58.6	51.44	63.8	89.01	60.7	161.20	54.3	272.88	40.3	110.26	54.8
	2. clothing	2.35	5.7	4.31	5.3	9.36	6.4	31.74	10.7	95.74	14.1	19.02	9.4
	3. house rent or maintenance	5.54	13.4	8.26	10.2	14.44	9.9	29.23	9.9	61.23	9.0	19.67	9.8
	4. fuel and light	3.88	9.4	5.84	7.2	9.60	6.6	15.36	5.2	25.39	3.7	11.21	5.6
	5. miscellaneous	5.35	12.9	10.82	13.5	24.08	16.4	59.15	19.9	223.21	32.9	41.14	20.4
	6. total	41.33	100.0	80.67	100.0	146.49	100.0	296.68	100.0	678.45	100.0	201.30	100.0
	number of households	174		433		491		220		31		1349	
	average size of households	2.28		4.35		5.79		7.10		7.48		5.13	
3. all localities	1. food	25.59	70.9	53.22	69.0	88.98	63.1	158.48	55.0	274.69	40.9	84.93	59.4
	2. clothing	1.62	4.5	4.36	5.7	10.37	7.3	31.93	11.1	101.65	15.2	12.93	9.0
	3. house rent or maintenance	0.82	2.3	2.35	3.1	9.11	6.5	23.23	8.0	47.97	7.2	9.06	6.3
	4. fuel and light	3.39	9.4	6.03	7.8	8.78	6.2	14.56	5.0	24.19	3.6	8.50	5.9
	5. miscellaneous	4.66	12.9	11.08	14.4	23.87	16.9	60.44	20.9	222.12	33.1	27.80	19.4
	6. total	36.08	100.0	77.04	100.0	141.11	100.0	288.64	100.0	670.62	100.0	143.22	100.0

TABLE 12: AVERAGE MONTHLY RECEIPTS PER HOUSEHOLD (OVER ALL HOUSEHOLDS) DISTRIBUTED BY SOURCE OF RECEIPTS AND MONTHLY EXPENDITURE LEVELS: JULY—SEPTEMBER 1953

	government colonies								other localities								all localities							
		average monthly receipts per h.h. from								average monthly receipts per h.h. from								average monthly receipts per h.h. from						
monthly expenditure in Rs.	number of households	occupation	government doles	other sources	non-government loan	non-government help	own supply	total	number of households	occupation	government doles	other sources	non-government loan	non-government help	own supply	total	number of households	occupation	government doles	other sources	non-government loan	non-government help	own supply	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)
1 – 25	39	7.5	8.8	1.9	0.5	2.8	0.1	21.6	2	7.5	–	–	–	–	–	7.5	41	7.5	8.4	1.8	0.5	2.7	0.1	21.0
26 – 50	115	22.3	5.9	8.7	11.0	2.6	–	50.5	18	37.5	0.3	6.6	–	12.0	–	56.4	133	24.4	5.1	8.4	9.5	3.9	–	51.3
51 – 75	174	44.0	4.0	9.8	8.2	2.6	0.2	68.8	30	55.2	–	5.2	4.3	3.4	1.3	69.4	204	45.6	3.4	9.2	7.6	2.8	0.4	69.0
76–100	183	59.8	4.6	14.0	16.0	4.0	0.3	98.7	46	81.9	–	6.2	4.6	4.1	1.3	98.1	229	64.2	3.7	12.4	13.7	4.0	0.5	98.5
101–150	208	89.7	2.1	21.6	16.9	5.9	0.6	136.8	124	104.0	0.8	22.6	7.3	3.1	0.5	138.3	332	95.0	1.6	22.0	13.3	4.9	0.6	137.4
151–200	70	121.4	4.0	31.5	18.8	2.0	0.8	178.5	89	142.3	0.3	36.6	7.9	12.2	1.4	200.7	159	133.2	1.9	34.3	12.7	7.7	1.1	190.9
201–250	36	140.5	1.1	35.5	36.7	12.5	–	226.3	59	171.7	4.8	48.2	12.4	5.1	0.3	242.5	95	159.9	3.4	43.3	21.6	7.9	0.2	236.3
251–500	47	237.0	–	87.1	27.2	10.9	1.1	363.3	78	248.1	3.2	80.2	16.0	35.6	3.3	386.4	125	244.0	2.0	82.8	20.2	26.3	2.4	377.7
501–750	7	445.0	–	103.6	21.4	22.9	–	592.9	17	477.5	–	145.3	–	5.9	–	628.7	24	468.1	–	133.2	6.2	10.8	–	618.3
751 & above	1	1350.0	–	–	–	–	–	1350.0	6	949.9	–	340.8	–	–	–	1290.7	7	1007.0	–	292.1	–	–	–	1299.1
total	880	78.6	3.8	20.6	15.0	4.7	0.4	123.1	469	159.9	1.4	43.1	8.4	11.0	1.2	225.0	1349	106.9	2.9	28.4	12.7	6.9	0.7	158.5

TABLE 12 (*contd.*): AVERAGE MONTHLY RECEIPTS PER HOUSEHOLD (OVER ALL HOUSEHOLDS) DISTRIBUTED BY SOURCE OF RECEIPTS AND MONTHLY EXPENDITURE LEVELS: JULY—SEPTEMBER 1953

	government colonies								other localities								all localities							
		average monthly receipts per h.h. from								average monthly receipts per h.h. from								average monthly receipts per h.h. from						
monthly expenditure in Rs.	number of households	occupation	government doles	other sources	non-government loan	non-govern-ment help	own supply	total	number of households	occupation	government doles	other sources	non-govern-ment loan	non-govern-ment help	own supply	total	number of households	occupation	government doles	other sources	non-govern-ment loan	non-govern-ment help	own supply	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)
percentage distribution of receipts by the different source of receipts and by expenditure levels																								
1– 25	39	34.7	40.7	8.8	2.3	13.0	0.5	100.0	2	100.0	–	–	–	–	–	100.0	41	35.7	39.9	8.6	2.4	12.9	0.5	100.0
26– 50	115	44.2	11.7	17.2	21.8	5.1	–	100.0	18	66.5	0.5	11.7	–	21.3	–	100.0	133	47.6	9.9	16.4	18.5	7.6	–	100.0
51– 75	174	64.0	5.8	14.2	11.9	3.8	0.3	100.0	30	79.5	–	7.5	6.2	4.9	1.9	100.0	204	66.1	4.9	13.3	11.0	4.1	0.6	100.0
76–100	183	60.5	4.7	14.2	16.2	4.1	0.3	100.0	46	83.5	–	6.3	4.7	4.2	1.3	100.0	229	65.1	3.8	12.6	13.9	4.0	0.6	100.0
101–150	208	65.6	1.5	15.8	12.4	4.3	0.4	100.0	124	75.2	0.6	16.3	5.3	2.2	0.4	100.0	332	69.1	1.2	16.0	9.7	3.6	0.4	100.0
151–200	70	68.1	2.2	17.7	10.5	1.1	0.4	100.0	89	71.0	0.1	18.2	3.9	6.1	0.7	100.0	159	69.7	1.0	18.0	6.7	4.0	0.6	100.0
201–250	36	62.1	0.5	15.7	16.2	5.5	–	100.0	59	70.8	2.0	19.9	5.1	2.1	0.1	100.0	95	67.8	1.4	18.3	9.1	3.3	0.1	100.0
251–500	47	65.2	–	24.0	7.5	3.0	0.3	100.0	78	64.2	0.8	20.8	4.1	9.2	0.9	100.0	125	64.7	0.5	21.9	5.3	7.0	0.6	100.0
501–750	7	75.0	–	17.5	3.6	3.9	–	100.0	17	76.0	–	23.1	–	0.9	–	100.0	24	75.8	–	21.5	1.0	1.7	–	1.000
751 & above	1	100.0	–	–	–	–	–	100.0	6	73.6	–	26.4	–	–	–	100.0	7	77.5	–	22.5	–	–	–	100.0
total	880	63.9	3.1	16.7	12.2	3.8	0.3	100.0	469	71.1	0.6	19.2	3.7	4.9	0.5	100.0	1349	67.5	1.8	17.9	8.0	4.4	0.4	100.0

TABLE 13 : NUMBER OF HOUSEHOLDS (EFFECTIVE SAMPLES) AND AVERAGE MONTHLY RECEIPTS BY SOURCES OF RECEIPTS AND EXPENDITURE LEVELS : JULY—SEPTEMBER 1953

monthly expenditure (Rs.)	number of total households	average total receipts	sources of receipts											
			occupation		government doles		other sources		non-government loan		non-government help		own supply	
			number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
						government colonies								
1 – 25	39	21.6	14	20.9	18	18.9	7	10.7	3	7.0	7	15.9	1	5.0
26 – 50	115	50.5	79	32.5	23	29.2	41	24.4	20	63.1	12	25.2	–	–
51 – 75	174	68.8	142	53.8	17	41.0	66	26.0	40	35.7	13	35.4	6	7.1
76–100	183	98.7	154	71.0	16	53.1	65	39.3	54	54.3	17	42.6	5	12.8
101–150	208	136.8	182	102.5	11	38.8	85	53.0	54	65.1	20	61.7	8	16.6
151–200	70	178.5	63	134.9	4	70.0	34	64.9	16	82.2	2	70.0	2	26.5
201–250	36	226.3	31	163.2	1	40.0	16	79.9	14	94.4	2	225.0	–	–
251–500	47	363.3	38	293.1	–	–	23	178.1	8	159.8	5	102.0	1	50.0
501–750	7	592.9	7	445.0	–	–	4	181.2	1	150.0	1	160.0	–	–
751 & above	1	1350.0	1	1350.0	–	–	–	–	–	–	–	–	–	–
total	880	123.1	711	97.4	90	36.7	341	53.2	210	63.0	79	51.8	23	15.1

TABLE 13 (*contd*) : NUMBER OF HOUSEHOLDS (EFFECTIVE SAMPLES) AND AVERAGE MONTHLY RECEIPTS BY SOURCES OF RECEIPTS AND EXPENDITURE LEVELS : JULY—SEPTEMBER 1953

			sources of receipts											
			occupation		government doles		other sources		non-government loan		non-government help		own supply	
monthly expenditure (Rs.)	number of total households	average total receipts	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
other localities														
1 – 25	2	7.5	2	7.5	–	–	–	–	–	–	–	–	–	–
26 – 50	18	56.4	12	56.3	1	5.0	4	29.5	–	–	6	36.2	–	–
51 – 75	30	69.4	27	61.3	–	–	9	17.4	5	25.6	4	25.5	1	40.0
76–100	46	98.1	42	89.8	–	–	14	20.3	5	42.0	5	37.8	5	11.6
101–150	124	138.3	116	111.1	5	21.0	46	60.6	14	65.1	10	38.7	8	8.5
151–200	89	200.7	76	166.6	1	30.0	38	85.6	7	100.7	10	108.5	1	125.0
201–250	59	242.5	52	194.8	3	94.3	27	105.3	6	122.0	2	150.0	1	15.0
251–500	78	386.4	62	312.1	1	250.0	43	145.5	6	208.3	12	231.4	4	64.0
501–750	17	628.7	16	507.4	–	–	12	205.8	–	–	1	100.0	–	–
751 & above	6	1290.7	4	1424.8	–	–	3	681.7	–	–	–	–	–	–
total	469	225.0	409	183.3	11	61.1	196	103.1	43	91.5	50	103.1	20	28.1

TABLE 13 (*contd.*) : NUMBER OF HOUSEHOLDS (EFFECTIVE SAMPLES) AND AVERAGE MONTHLY RECEIPTS BY SOURCES OF RECEIPTS AND EXPENDITURE LEVELS : JULY—SEPTEMBER 1953

monthly expenditure (Rs.)	number of total households	average total receipts	sources of receipt											
			occupation		government doles		other sources		non-government loan		non-government help		own supply	
			number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.	number of households	amount in Rs.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
all localities														
1 – 25	41	21.0	16	19.2	18	18.7	7	10.7	3	7.0	7	15.9	1	5.0
26 – 50	133	51.3	91	35.6	24	28.2	45	24.9	20	63.1	18	28.9	–	–
51 – 75	204	69.0	169	55.0	17	41.0	75	25.0	45	34.6	17	33.1	7	11.8
76–100	229	98.5	196	75.0	16	53.	79	35.9	59	53.3	22	41.5	10	12.2
101–150	332	137.4	298	105.8	16	33.2	131	55.7	68	65.1	30	54.0	16	12.6
151–200	159	190.9	139	152.2	5	62.0	72	75.8	23	87.8	12	102.1	3	59.3
201–250	95	236.3	83	183.0	4	80.7	34	95.8	20	102.7	4	187.5	1	15.0
251–500	125	377.7	100	304.9	1	250.0	66	156.9	14	180.6	17	193.3	5	61.2
501–750	24	618.3	23	488.4	–	–	16	199.6	1	150.0	2	130.0	–	–
751 & above	7	1299.1	5	1409.8	–	–	3	681.7	–	–	–	–	–	–
total	1349	158.5	1120	128.8	101	39.4	537	71.4	253	67.8	129	71.7	43	21.1

TABLE 14 : DISTRIBUTION OF SAMPLE HOUSEHOLDS BY LEVEL OF TOTAL MONTHLY RECEIPTS : JULY—SEPTEMBER 1953

total monthly receipts in Rs.	government colonies			other localities			all localities		
	number of households	percentage	cumulative percentage	number of households	percentage	cumulative percentage	number of households	percentage	cumulative percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1 – 25	36	4.1	4.1	3	0.6	0.6	39	2.9	2.9
26 – 50	116	13.2	17.3	19	4.1	4.7	135	10.0	12.9
51 – 75	154	17.5	34.8	25	5.3	10.0	179	13.3	26.2
76–100	180	20.4	55.2	45	9.6	19.6	225	16.7	42.9
101–150	200	22.7	77.9	105	22.4	42.0	305	22.6	65.5
151–200	87	9.9	87.8	93	19.8	61.8	180	13.3	78.8
201–250	41	4.7	92.5	59	12.6	74.4	100	7.4	86.2
251–500	52	5.9	98.4	91	19.4	93.8	143	10.6	96.8
501–750	9	1.0	99.4	19	4.1	97.9	28	2.1	98.9
751 & above	5	0.6	100.0	10	2.1	100.0	5	1.1	100.0
total	880	100.0	—	469	100.0	—	1349	100.0	—

TABLE 15 : DISTRIBUTION OF HOUSEHOLDS BY PERCENTAGE OF INCOME (FROM OCCUPATION, "HOME SUPPLY" AND "OTHER SOURCES") TO TOTAL EXPENDITURE : JULY—SEPTEMBER 1953

percentage of income to expenditure	government colonies				other localities				all localities			
	number of households	per-cent-age	cumulative percentage		number of households	per-cent-age	cumulative percentage		number of households	per-cent-age	cumulative percentage	
			up-wards	down-wards			up-wards	down-wards			up-wards	down-wards
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
nil	109	12.4	100.0	12.4	19	4.1	100.0	4.1	128	9.5	100.0	9.5
upto 20	24	2.7	87.6	15.1	8	1.7	95.9	5.8	32	2.4	90.5	11.9
21 – 40	40	4.6	84.9	19.7	14	3.0	94.2	8.8	54	4.0	88.1	15.9
41 – 60	72	8.2	80.3	27.9	23	4.9	91.2	13.7	95	7.0	84.1	22.9
61 – 80	91	10.3	72.1	38.2	38	8.1	86.3	21.8	129	9.6	77.1	32.5
81 – 90	42	4.8	61.8	43.0	15	3.2	78.2	25.0	57	4.2	67.5	36.7
91–100	183	20.8	57.0	63.8	134	28.6	75.0	53.6	317	23.5	63.3	60.2
above 100	319	36.2	36.2	100.0	218	46.4	46.4	100.0	537	39.8	39.8	100.0
total	880	100.0	—	—	469	100.0	—	—	1349	100.0	—	—

TABLE 16 : NUMBER OF HOUSEHOLDS USING NON-GOVERNMENT LOANS FOR CONSUMPTION IN THE PERIOD UNDER INVESTIGATION ARRANGED BY TYPES OF SECURITIES GIVEN WITH OR WITHOUT INTEREST : JULY—EPTEMBER 1953

type of security given	with interest		without interest		total	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
government colonies						
1. without security	5	2.4	46	21.9	51	24.3
2. personal	19	9.0	115	54.8	134	63.8
3. gold	3	1.4	10	4.8	13	6.2
4. others	1	0.5	—	—	1	0.5
5. not returned	1	0.5	10	4.7	11	5.2
6. total	29	13.8	181	86.2	210	100.0
other localities						
1. without security	2	4.7	24	55.8	26	60.5
2. personal	1	2.3	13	30.3	14	32.6
3. gold	—	0.0	1	2.3	1	2.3
4. others	1	2.3	—	—	1	2.3
5. not returned	—	—	1	2.3	1	2.3
6. total	4	9.3	39	90.7	43	100.0
all localities						
1. without security	7	2.7	70	27.8	77	30.5
2. personal	20	7.9	128	50.6	148	58.5
3. gold	3	1.2	11	4.3	14	5.5
4. others	2	0.8	—	0.0	2	0.8
5. not returned	1	0.4	11	4.3	12	4.7
6. total	33	13.0	220	87.0	253	100.0

TABLE 17 : NUMBER OF HOUSEHOLDS HAVING LOANS AND AMOUNT OF OUTSTANDING LOANS PER HOUSEHOLD : JULY—SEPTEMBER 1953

localities	number of households	households having outstanding loan	percentage to total	amount per household in Rs.		
				government	private	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. government colonies	880	562	64.0	460.8	675.5	1136.3
2. other localities	469	191	40.8	759.4	2075.9	2835.3
3. all localities	1349	753	55.9	536.5	1030.7	1567.2

TABLE 18 : DISTRIBUTION OF HOUSEHOLDS BY LEVELS OF OUTSTANDING (GOVERNMENT AND NON-GOVERNMENT) LOANS : JULY—SEPTEMBER 1953

levels of outstanding loan (Rs.)	government colonies			other localities			all localities		
	number of households	percentage	cumulative percentage	number of households	percentage	cumulative percentage	number of households	percentage	cumulative percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
nil	318	36.0	36.0	278	59.2	59.2	596	44.1	44.1
1 – 99	35	4.0	40.0	9	1.9	61.1	44	3.3	47.4
100 – 199	26	3.0	43.0	6	1.3	62.4	32	2.4	49.8
200 – 299	41	4.7	47.7	13	2.8	65.2	54	4.0	53.8
300 – 399	49	5.6	53.3	12	2.6	67.8	61	4.5	58.3
400 – 499	34	3.9	57.2	10	2.1	69.9	44	3.3	61.6
500 – 999	182	20.7	77.9	44	9.4	79.3	226	16.7	78.3
1000 – 1499	85	9.7	87.6	29	6.2	85.5	114	8.5	86.8
1500 – 1999	38	4.3	91.9	9	1.9	87.4	47	3.5	90.3
2000 – 2999	35	4.0	95.9	19	4.1	91.5	54	4.0	94.3
3000 – 3999	11	1.2	97.1	6	1.3	92.8	17	1.3	95.6
4000 – 4999	6	0.7	97.8	9	1.9	94.7	15	1.1	96.7
5000 – 9999	18	2.0	99.8	12	2.6	97.3	30	2.2	98.9
10000–14999	1	0.1	99.9	8	1.7	99.0	9	0.7	99.6
15000–19999	—	—	—	3	0.6	99.6	3	0.2	99.8
20000 & above	1	0.1	100.0	2	0.4	100.0	3	0.2	100.0
total	880	100.0	—	469	100.0	—	1349	100.0	—

TABLE 19 : NUMBER OF NON-GOVERNMENT LOANS OBTAINED SINCE MIGRATION WITH OR WITHOUT INTEREST ARRANGED BY TYPES OF SECURITIES GIVEN : JULY—SEPTEMBER 1953

type of security given	with interest		without interest		total	
	number	percentage	number	percentage	number	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			government colonies (loans of 438 households)			
1. without security	32	3.3	220	22.8	252	26.1
2. personal	95	9.9	493	51.1	588	61.0
3. gold	18	1.9	50	5.2	68	7.0
4. land and house	—	—	—	—	—	—
5. others	12	1.2	4	0.4	16	1.7
6. not returned	—	—	41	4.2	41	4.2
7. total	157	16.3	808	83.7	965	100.0
			other localities (loans of 164 households)			
1. without security	2	7.4	145	48.8	167	56.2
2. personal	16	5.4	79	26.6	95	32.0
3. gold	6	2.0	10	3.4	16	5.4
4. land and house	1	0.3	1	0.3	2	0.7
5. others	7	2.4	1	0.3	8	2.7
6. not returned	2	0.7	7	2.4	9	3.0
7. total	54	18.2	243	81.8	297	100.0
			all localities (loans of 602 households)			
1. without security	54	4.3	365	28.9	419	33.2
2. personal	111	8.7	572	45.3	683	54.0
3. gold	24	1.9	60	4.8	84	6.7
4. land and house	1	0.1	1	0.1	2	0.2
5. others	19	1.5	5	0.4	24	1.9
6. not returned	2	0.2	48	3.8	50	4.0
7. total	211	16.7	1051	83.3	1262	100.0

TABLE 20: NUMBER OF NON-GOVERNMENT LOANS OBTAINED SINCE MIGRATION WITH OR WITHOUT INTEREST BY PURPOSES FOR WHICH THESE WERE TAKEN AND BY TYPES OF SECURITIES GIVEN: JULY—SEPTEMBER 1953

type of security offered	purpose for taking loan (private)																				
	house-building			trade			profession			maintenance			others			not returned			all purpose		
	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
government colonies (loans of 438 households)																					
1. personal	4	–	4	13	37	50	–	–	–	68	436	504	9	20	29	1	–	1	95	493	588
2. gold	1	–	1	2	7	9	–	–	–	13	43	56	2	–	2	–	–	–	18	50	68
3. land and house	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
4. without security	–	–	–	13	41	54	2	1	3	14	168	182	3	10	13	–	–	–	32	220	252
5. others	1	–	1	5	–	5	–	–	–	6	4	10	–	–	–	–	–	–	12	4	16
6. not returned	–	–	–	–	10	10	–	–	–	–	24	24	–	5	5	–	2	2	–	41	41
7. total	6	–	6	33	95	128	2	1	3	101	675	776	14	35	49	1	2	3	157	808	965

TABLE 20 (*contd.*) : NUMBER OF NON-GOVERNMENT LOANS OBTAINED SINCE MIGRATION WITH OR WITHOUT INTEREST BY PURPOSES FOR WHICH THESE WERE TAKEN AND BY TYPES OF SECURITIES GIVEN : JULY—SEPTEMBER 1953

type of security offered	purpose for taking loan (private)																		all purpose		
	house-building			trade			profession			maintenance			others			not returned					
	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
other localities (loans of 164 households)																					
1. personal	–	–	–	11	26	37	–	–	–	3	52	55	2	1	3	–	–	–	16	79	95
2. gold	–	–	–	–	1	1	–	–	–	6	8	14	–	1	1	–	–	–	6	10	16
3. land and house	–	–	–	1	–	1	–	–	–	–	–	–	–	1	1	–	–	–	1	1	2
4. without security	–	–	–	8	30	38	–	–	–	13	100	113	1	14	15	–	1	1	22	145	167
5. others	–	–	–	5	1	6	–	–	–	1	–	1	1	–	1	–	–	–	7	1	8
6. not returned	–	–	–	2	5	7	–	–	–	–	2	2	–	–	–	–	–	–	2	7	9
7. total	–	–	–	27	63	90	–	–	–	23	162	185	4	17	21	–	1	1	54	243	297

TABLE 20 (*contd.*) : NUMBER OF NON-GOVERNMENT LOANS OBTAINED SINCE MIGRATION WITH OR WITHOUT INTEREST BY PURPOSES FOR WHICH THESE WERE TAKEN AND BY TYPES OF SECURITIES GIVEN : JULY—SEPTEMBER 1953

type of security offered	purpose for taking loan (private)																				
	house-building			trade			profession			maintenance			others			not returned			all purpose		
	with interest	without interest	total	with interes	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total	with interest	without interest	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
all localities (loans of 602 households)																					
1. personal	4	–	4	24	63	87	–	–	–	71	488	559	12	21	33	–	–	–	111	572	683
2. gold	1	–	1	2	8	10	–	–	–	19	51	70	2	1	3	–	–	–	24	60	84
3. land and house	–	–	–	1	–	1	–	–	–	–	–	–	–	1	1	–	–	–	1	1	2
4. without security	–	–	–	21	71	92	2	1	3	27	268	295	4	24	28	–	1	1	54	365	419
5. others	1	–	1	10	1	11	–	–	–	7	4	11	1	–	1	–	–	–	19	5	24
6. not returned	–	–	–	2	15	17	–	–	–	–	26	26	–	5	5	–	2	2	2	48	50
7. *total*	6	–	6	60	158	218	2	1	3	124	837	961	19	52	71	–	3	3	211	1051	1262

TABLE 21 : AMOUNT OF GOVERNMENT LOANS AND GRANTS RECEIVED PER HOUSEHOLD AND THE EXTENT OF THEIR UTILISATION; AMOUNT OF INTEREST PAID AND LOANS REPAID PER HOUSEHOLD : JULY—SEPTEMBER 1953

type of loans and grants	number of house-hold	receipts			utilisation						percentage utilised for the purpose	balance in hand in Rs.	loan repaid to date in Rs.	interest paid to date in Rs.
		average time interval in months	average number of instalments	amount per household in Rs.	for the purpose in Rs.	for other productive purposes in Rs.	for consumption in Rs.	as loan to others in Rs.	as help to others in Rs.	for repayment of private loans in Rs.				
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
					government colonies									
1. house-building	3	1.00	1.33	4750.0	4500.0	—	—	—	—	—	94.7	250.0	0.0	0.0
2. trade	284	0.35	1.05	796.4	598.2	15.0	158.3	3.3	0.9	3.9	75.1	16.8	8.7	4.6
3. profession	25	1.74	1.08	619.0	454.6	26.0	138.4	—	—	—	73.4	0.0	0.2	0.0
4. maintenance and education loan	12	1.62	1.75	563.3	563.3	—	—	—	—	—	100.0	0.0	101.6	0.0
5. maintenance grant	4	0.96	20.25	787.5	437.5	—	—	—	—	350.0	55.6	0.0	0.0	0.0
6. education grant	—	—	—	—	—	—	—	—	—	—	—	—	—	—
					other localities									
1. house-building	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2. trade	66	1.71	1.24	2117.4	1741.3	34.1	129.9	—	—	—	82.2	212.1	4.5	18.8
3. profession	4	9.20	1.25	950.0	812.5	50.0	87.5	—	—	—	85.5	0.0	0.0	0.0
4. maintenance	2	0.00	1.00	550.0	550.0	—	—	—	—	—	100.0	0.0	105.0	0.0
5. education	1	0.00	2.00	1500.0	1500.0	—	—	—	—	—	100.0	0.0	600.0	0.0
6. maintenance grant	4	2.67	3.00	123.5	87.5	—	—	—	—	—	70.9	36.0	0.0	0.0
7. education grant	2	1.08	6.00	133.0	133.0	—	—	—	—	—	100.0	0.0	0.0	0.0

TABLE 21 (*contd.*) : AMOUNT OF GOVERNMENT LOANS AND GRANTS RECEIVED PER HOUSEHOLD AND THE EXTENT OF THEIR UTILISATION; AMOUNT OF INTEREST PAID AND LOANS REPAID PER HOUSEHOLD : JULY—SEPTEMBER 1953

type of loans and grants	number of households	receipts			utilisation						percentage utilised for the purpose	balance in hand in Rs.	loan repaid to date in Rs.	interest paid to date in Rs.
		average time interval in months	average number of instalments	amount per household in Rs.	for the purpose in Rs.	for other productive purposes in Rs.	for consumption in Rs.	as loan to others in Rs.	as help to others in Rs.	for repayment of private loans in Rs.				
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
							all localities							
1. house-building	3	1.00	1.33	4750.0	4500.0	—	—	—	—	—	94.7	250.0	0.0	0.0
2. trade	350	0.65	1.08	1045.5	813.7	18.6	153.0	2.7	0.7	3.1	77.8	53.7	7.9	7.3
3. profession	29	2.91	1.10	664.7	504.0	29.3	131.4	—	—	—	75.8	0.0	0.2	0.0
4. maintenance	14	1.48	1.64	561.4	561.4	—	—	—	—	—	100.0	0.0	102.1	0.0
5. education	1	0.00	2.00	1500.0	1500.0	—	—	—	—	—	100.0	0.0	600.0	0.0
6. maintenance grant	8	1.18	11.62	455.5	262.5	—	—	—	—	—	57.6	18.0	0.0	0.0
7. education grant	2	1.08	6.00	133.0	133.0	—	—	—	—	—	100.0	0.0	0.0	0.0

TABLE 22 : DISTRIBUTION OF HOUSEHOLDS BY TOTAL AMOUNTS OF GOVERNMENT LOANS : JULY—SEPTEMBER 1953

total amount of loan received (Rs.)	government colonies		other localities		all localities	
	number of house-holds	percent-age	number of house-holds	percent-age	number of house-holds	percent-age
(1)	(2)	(3)	(4)	(5)	(6)	(7)
nil	559	63.6	396	84.4	955	70.9
1 - 99	–	–	–	–	–	–
100 - 199	9	1.0	1	0.2	10	0.7
200 - 299	24	2.7	5	1.1	29	2.1
300 - 399	44	5.0	7	1.5	51	3.8
400 - 499	32	3.6	5	1.1	37	2.7
500 - 999	150	17.1	23	4.9	172	12.8
1000 - 2999	52	5.9	20	4.3	72	5.4
3000 - 4999	1	0.1	2	0.4	3	0.2
5000 - 9999	8	0.9	8	1.7	16	1.2
10000–14999	1	0.1	1	0.2	2	0.1
15000–19999	–	–	–	–	–	–
20000 & above	–	–	1	0.2	1	0.1
total	880	100.0	469	100.0	1349	100.0

TABLE 23: DISTRIBUTION OF HOUSEHOLDS BY TYPES AND AMOUNTS OF GOVERNMENT LOANS: JULY—SEPTEMBER 1953

amount of loan in Rs.	number of households receiving loans for						
	house-building	trade	profession	maintenance	education	all types	
						number	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			government colonies				
nil	877	596	855	868	880	559	63.6
1 – 99	–	–	–	–	–	–	0.0
100 – 199	–	6		1	–	9	1.0
200 – 299	–	20	3	2	–	24	2.7
300 – 399	–	40	5	1	–	44	5.0
400 – 499	–	29	1	1	–	32	3.6
500 – 999	1	135	11	5	–	150	17.1
1000–1499	–	33	–	2	–	36	4.1
1500–1999	–	6	1	–	–	7	0.8
2000–2499	–	6	2	–	–	8	0.9
2500–4999	1	1	–	–	–	2	0.2
5000–9999	1	7	–	–	–	8	0.9
10000 & above	–	1	–	–	–	1	0.1
total	880	880	880	880	880	880	100.0
			other localities				
nil	469	403	465	467	468	396	84.4
1 – 99	–	–	–	–	–	–	–
100 – 199	–	1	–	–	–	1	0.2
200 – 299	–	5	–	–	–	5	1.1
300 – 399	–	7	–	–	–	7	1.5
400 – 499	–	5	–	–	–	5	1.1
500 – 999	–	19	2	2	–	23	4.9
1000–1499	–	9	1	–	–	10	2.2
1500–1999	–	4	1	–	1	6	1.3
2000–2499	–	3	–	–	–	3	0.6
2500–4999	–	3	–	–	–	3	0.6
5000–9999	–	8	–	–	–	8	1.7
10000 & above	–	2	–	–	–	2	0.4
total	469	469	469	469	469	469	100.0

TABLE 23 (*contd.*) : DISTRIBUTION OF HOUSEHOLDS BY TYPES AND AMOUNTS OF GOVERNMEMT LOANS : JULY—SEPTEMBER 1953

amount of loan in Rs.	number of households receiving loans for						
	house-building	trade	profession	maintenance	education loan	all types	
						number	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
all localities							
nil	1346	999	1320	1335	1348	955	70.8
1 – 99	–	–	–	–	–	–	–
100 – 199	–	7	2	1	–	10	0.7
200 – 299	–	25	3	2	–	29	2.1
300 – 399	–	47	5	1	–	51	3.8
400 – 499	–	34	1	1	–	37	2.7
500 – 999	1	154	13	7	–	173	12.8
1000–1499	–	42	1	2	–	46	3.5
1500–1999	–	10	2	–	1	13	1.0
2000–2499	–	9	2	–	–	11	0.8
2500–4999	1	4	–	–	–	5	0.4
5000–9999	1	15	–	–	–	16	1.2
10000 & above	–	3	–	–	–	3	0.2
total	1349	1349	1349	1349	1349	1349	100.0

TABLE 24: DISTRIBUTION OF HOUSEHOLDS BY LEVELS OF UTILISATION IN PURPOSES FOR WHICH THE LOANS WERE GRANTED: JULY—SEPTEMBER 1953

levels of utilisation expressed in percentage	government colonies						other localities						all localities					
	number of households received loans			percentage			number of households received loans			percentage			number of households received loans			percentage		
	trade	profe-ssion	all types	trade	profe-ssion	all types	trade	profe-ssion	all types	trade	profe-ssion	all types	trade	profe-ssion	all types	trade	profe-ssion	all types
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
upto 10.0	61	8	69	21.4	32.0	21.5	11	-	10	16.7	-	13.7	72	8	79	20.6	27.6	20.1
10.1 - 20.0	5	-	6	1.8	-	1.9	-	-	-	-	-	-	5	-	6	1.4	-	1.5
20.1 - 30.0	7	-	8	2.5	-	2.5	-	-	1	-	-	1.4	7	-	9	2.0	-	2.3
30.1 - 40.0	3	1	4	1.1	4.0	1.2	1	1	2	1.5	25.0	2.7	4	2	6	1.1	6.9	1.5
40.1 - 50.0	5	-	5	1.8	-	1.6	-	-	-	-	-	-	5	-	5	1.4	-	1.3
50.1 - 60.0	15	2	16	5.3	8.0	5.0	1	-	1	1.5	-	1.4	16	2	17	4.6	6.9	4.3
60.1 - 70.0	28	2	30	9.9	8.0	9.3	3	-	3	4.5	-	4.1	31	2	33	8.9	6.9	8.4
70.1 - 80.0	10	1	11	3.5	4.0	3.4	5	-	5	7.6	-	6.8	15	1	16	4.3	3.4	4.1
80.1 - 90.0	8	-	8	2.8	-	2.5	1	1	2	1.5	25.0	2.7	9	1	10	2.6	3.4	2.5
90.1-100.0	142	11	164	49.9	44.0	51.1	44	2	49	66.7	50.0	67.2	186	13	213	53.1	44.9	54.0
total	284	25	321	100.0	100.0	100.0	66	4	73	100.0	100.0	100.0	350	29	394	100.0	100.0	100.0

N.B. Distribution of households who received loans for house-building and other less important items has not been shown separately because their number is very small.

TABLE 25 : DISTRIBUTION OF HOUSEHOLDS BY THE NATURE OF USE OF LAND UNDER OCCUPATION : JULY—SEPTEMBER 1953

use of land	number of households occupying land			percentage		
	government colonies	other localities	all localities	government colonies	other localities	all localities
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. cultivation	3	1	4	0.4	0.3	0.3
2. housing	783	273	1056	96.2	79.6	91.3
3. others	1	14	15	0.1	4.1	1.3
4. cultivation and housing	2	–	2	0.2	–	0.2
5. housing and others	25	55	80	3.1	16.0	6.9
6. sub-total	814	343	1157	100.0	100.0	100.0
7. no information	66	126	192	–	–	–
8. total	880	469	1349	–	–	–

TABLE 26 : DISTRIBUTION OF HOUSEHOLDS BY NUMBER OF ROOMS AND TYPE OF POSSESSION OF LIVING ACCOMMODATION : JULY—SEPTEMBER 1953

number of rooms	rented		owned		not returned		total	
	number of house-holds	percen-tage	number of house-holds	percen-tage	number of house-holds	percen-tage	number of house-holds	percen-tage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				government colonies				
1	750	86.4	3	30.0	–	–	753	85.7
2	87	10.0	3	30.0	1	100.0	91	10.3
3	20	2.3	3	30.0	–	–	23	2.6
4	7	0.8	–	–	–	–	7	0.8
5	2	0.2	–	–	–	–	2	0.2
6 & over	2	0.2	1	10.0	–	–	3	0.3
not returned	1	0.1	–	–	–	–	1	0.1
total	869	100.0	10	100.0	1	100.0	880	100.0
include percentage	(98.8)	–	(1.1)	–	(0.1)	–	(100.0)	–
				other localities				
1	176	43.0	12	24.5	4	36.4	192	40.9
2	148	36.2	21	42.9	5	45.4	174	37.1
3	53	13.0	12	24.5	1	9.1	66	14.1
4	25	6.1	3	6.1	–	–	28	6.0
5	2	0.5	–	–	–	–	2	0.4
6 & over	5	1.2	1	2.0	–	–	6	1.3
not returned	–	–	–	–	1	9.1	1	0.2
total	409	100.0	49	100.0	11	100.0	469	100.0
include percentage	(7.82)	–	(10.4)	–	(2.4)	–	(100.0)	–
				all localities				
1	926	72.5	15	25.4	5	41.7	946	70.1
2	235	18.4	24	40.7	5	41.7	264	19.6
3	73	5.7	15	25.4	1	8.3	89	6.6
4	32	2.5	3	5.1	–	–	35	2.6
5	4	0.3	–	–	–	–	4	0.3
6 & over	7	0.5	2	3.4	–	–	9	0.7
not returned	1	0.1	–	–	1	8.3	2	0.1
total	1278	100.0	59	100.0	12	100.0	1349	100.0
include percentage	(94.7)	–	(4.4)	–	(0.9)	–	(100.0)	–

NATIONAL SAMPLE SURVEY : SURVEY OF DISPLACED PERSONS IN BOMBAY

TABLE 27 : PERCENTAGE DISTRIBUTION OF HOUSEHOLDS BY FLOOR SPACE AND TYPE OF CONSTRUCTION : JULY—SEPTEMBER 1953

base : households under each level of floor space

level of floor space in sq. ft. (1)	totally brick built (2)	mixed (3)	totally non-brick built (4)	total (5)	number of households (6)
		government colonies			
upto 100	4.3	80.0	15.7	100.0	70
101 — 200	7.3	80.5	12.2	100.0	468
201 — 300	14.7	79.8	5.5	100.0	218
301 — 400	15.6	79.7	4.7	100.0	64
above 400	3.3	86.7	10.0	100.0	60
total	9.2	80.7	10.1	100.0	880
		other localities			
upto 100	21.2	27.5	51.3	100.0	80
101 — 200	27.4	38.4	34.2	100.0	146
201 — 300	29.5	27.7	42.8	100.0	112
301 — 400	37.3	35.6	27.1	100.0	59
above 400	34.7	36.1	29.2	100.0	72
total	29.2	33.3	37.5	100.0	469
		all localities			
upto 100	13.3	52.0	34.7	100.0	150
101 — 200	12.1	70.5	17.4	100.0	614
201 — 300	19.7	62.1	18.2	100.0	330
301 — 400	26.0	58.6	15.4	100.0	123
above 400	20.5	59.1	20.4	100.0	132
total	16.2	64.2	19.6	100.0	1349

TABLE 28 : PERCENTAGE DISTRIBUTION OF HOUSEHOLDS BY FLOOR SPACE AND TYPE OF CONSTRUCTION : JULY—SEPTEMBER 1953

base : households under each type of construction

level of floor space in sq. ft.	totally brick built	mixed	totally non-brick built	total
(1)	(2)	(3)	(4)	(5)
		government colonies		
upto 100	3.7	7.9	12.4	7.9
101 — 200	42.0	53.1	64.0	53.2
201 — 300	39.6	24.5	13.5	24.8
301 — 400	12.3	7.2	3.4	7.3
above 400	2.4	7.3	6.7	6.8
total	100.0	100.0	100.0	100.0
number of households	81	710	89	880
		other localities		
upto 100	12.4	14.1	23.3	17.0
101 — 200	29.2	35.8	28.4	31.1
201 — 300	24.1	20.0	27.3	23.9
301 — 400	16.1	13.5	9.1	12.6
above 400	18.2	16.6	11.9	15.4
total	100.0	100.0	100.0	100.0
number of households	137	156	176	469
		all localities		
upto 100	9.2	9.0	19.6	11.1
101 — 200	34.0	50.0	40.4	45.5
201 — 300	29.7	23.7	22.6	24.5
301 — 400	14.7	8.3	7.2	9.1
above 400	12.4	9.0	10.2	9.8
total	100.0	100.0	100.0	100.0
number of households	218	866	265	1349

TABLE 29 : DISTRIBUTION OF HOUSEHOLDS BY FLOOR SPACE AND SIZE OF HOUSEHOLDS WITH PERCENTAGE DISTRIBUTION BY FLOOR SPACE : JULY—SEPTEMBER 1953

floor space in sq. ft.	persons per household													
	1		2		3		4		5		above 5		total	
	number of h.h.	percentage	number of h.h.	percentage	number of h.h.	percentage	number of h.h.	percentage	number of h.h.	percentage	number of h.h.	percentage	number of h.h.	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
government colonies														
upto 100	15	23.8	20	22.7	8	5.6	5	3.8	8	6.8	14	4.2	70	7.9
101–200	31	49.3	48	54.6	71	49.7	75	56.3	62	53.1	181	53.8	468	53.2
201–300	12	19.0	13	14.8	49	34.2	33	24.8	28	23.9	83	24.7	218	24.8
301–400	4	6.3	4	4.5	9	6.3	14	10.5	10	8.5	23	6.8	64	7.3
401 & above	1	1.6	3	3.4	6	4.2	6	4.6	9	7.7	35	10.5	60	6.8
total	63	100.0	88	100.0	143	100.0	133	100.0	117	100.0	336	100.0	880	100.0
other localities														
upto 100	8	30.7	12	30.0	9	16.7	17	27.8	11	15.1	23	10.7	80	17.0
101–200	10	38.6	13	32.5	17	31.4	23	37.7	23	31.5	60	27.9	146	31.1
201–300	4	15.4	10	25.0	15	27.7	12	19.7	17	23.3	54	25.1	112	23.9
301–400	1	3.8	2	5.0	7	13.0	3	4.9	8	11.0	38	17.7	59	12.6
401 & above	3	11.5	3	7.5	6	11.2	6	9.9	14	19.1	40	18.6	72	15.4
total	26	100.0	40	100.0	54	100.0	61	100.0	73	100.0	215	100.0	469	100.0
all localities														
upto 100	23	25.9	32	25.0	17	8.6	22	11.3	19	10.0	37	6.7	150	11.1
101–200	41	46.1	61	47.7	88	44.7	98	50.5	85	44.8	241	43.7	614	45.5
201–300	16	18.0	23	18.0	64	32.5	45	23.2	45	23.6	137	24.9	330	24.5
301–400	5	5.6	6	4.7	16	8.1	17	8.8	18	9.5	61	11.1	123	9.1
401 & above	4	4.4	6	4.6	12	6.1	12	6.2	23	12.1	75	13.6	132	9.8
total	89	100.0	128	100.0	197	100.0	194	100.0	190	100.0	551	100.0	1349	100.0

TABLE 30 : DISTRIBUTION OF HOUSEHOLDS BY FLOOR SPACE AND SIZE OF HOUSEHOLDS WITH PERCENTAGE DISTRIBUTION BY SIZE OF HOUSEHOLDS : JULY—SEPTEMBER 1953

floor space in sq. ft.	government colonies							other localities							all localities						
	persons per household							persons per household							persons per household						
	1	2	3	4	5	above 5	total	1	2	3	4	5	above 5	total	1	2	3	4	5	above 5	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
upto 100	15	20	8	5	8	14	70	8	12	9	17	11	23	80	23	32	17	22	19	37	150
p.c.	(21.4)	(28.7)	(11.4)	(7.1)	(11.4)	(20.0)	(100.0)	(10.0)	(15.0)	(11.2)	(21.2)	(13.8)	(28.8)	(100.0)	(15.3)	(21.3)	(11.3)	(14.7)	(12.7)	(24.7)	(100.0)
101–200	31	48	71	75	62	181	468	10	13	17	23	23	60	146	41	61	88	98	85	241	614
p.c.	(6.6)	(10.3)	(15.2)	(16.0)	(13.2)	(38.7)	(100.0)	(6.8)	(8.9)	(11.6)	(15.8)	(15.8)	(41.1)	(100.0)	(6.7)	(9.9)	(14.3)	(16.0)	(13.8)	(39.3)	(100.0)
201–300	12	13	49	33	28	83	218	4	10	15	12	17	54	112	16	23	64	45	45	137	330
p.c.	(5.5)	(6.0)	(22.5)	(15.1)	(12.8)	(38.1)	(100.0)	(3.6)	(8.9)	(13.4)	(10.7)	(15.2)	(48.2)	(100.0)	(4.8)	(7.0)	(19.4)	(13.6)	(13.6)	(41.6)	(100.0)
301–400	4	4	9	14	10	23	64	1	2	7	3	8	38	59	5	6	16	17	18	61	123
p.c.	(6.2)	(6.2)	(14.1)	(21.9)	(15.6)	(36.0)	(100.0)	(1.7)	(3.4)	(11.9)	(5.1)	(13.6)	(64.3)	(100.0)	(4.1)	(4.9)	(13.0)	(13.8)	(14.6)	(49.6)	(100.0)
above 400	1	3	6	6	9	35	60	3	3	6	6	14	40	72	4	6	12	12	23	75	132
p.c.	(1.7)	(5.0)	(10.0)	(10.0)	(15.0)	(58.3)	(100.0)	(4.2)	(4.2)	(8.3)	(8.3)	(19.4)	(55.6)	(100.0)	(3.0)	(4.5)	(9.1)	(9.1)	(17.4)	(56.9)	(100.0)
total	63	88	143	133	117	336	880	26	40	54	61	73	215	469	89	128	197	194	190	551	1349
p.c.	(7.2)	(10.0)	(16.2)	(15.1)	(13.3)	(38.2)	(100.0)	(5.5)	(8.5)	(11.5)	(13.0)	(15.6)	(45.9)	(100.0)	(6.6)	(9.5)	(14.6)	(14.4)	(14.1)	(40.8)	(100.0)

TABLE 31 : DISTRIBUTION OF HOUSEHOLDS BY TYPE OF CONSTRUCTION OF PLINTH, WALL AND ROOF OF LIVING ACCOMMODATION : JULY—SEPTEMBER 1953

type of construction			government colonies		other localities		all localities	
plinth	wall	roof	number of house-holds	percentage	number of house-holds	percentage	number of house-holds	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
brick	brick	brick	81	9.2	137	29.2	218	16.2
,,	,,	tile	449	51.0	70	14.9	519	38.5
,,	,,	c.i. sheets	208	23.6	44	9.4	252	18.7
,,	,,	others	5	0.6	26	5.5	31	2.3
,,	c.i. sheets	c.i. sheets	11	1.3	8	1.7	19	1.4
mud	brick	tile	22	2.5	4	0.9	26	1.9
,,	,,	c.i. sheets	15	1.7	4	0.9	19	1.4
,,	mud	tile	12	1.3	3	0.6	15	1.1
others	others	others	77	8.8	173	36.9	250	18.5
total			880	100.0	469	100.0	1349	100.0

TABLE 32 : DISTRIBUTION OF HOUSEHOLDS BY TYPE OF LIGHTING : JULY—SEPTEMBER 1953

type of lighting	government colonies		other localities		all localities	
	number of house-holds	percentage	number of house-holds	percentage	number of house-holds	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. electric	326	37.0	334	71.2	660	48.9
2. kerosene	542	61.7	133	28.4	675	50.1
3. others	11	1.2	—	—	11	0.8
4. not recorded	1	0.1	2	0.4	3	0.2
5. all types	880	100.0	469	100.0	1349	100.0

TABLE 33 : DISTRIBUTION OF HOUSEHOLDS BY SOURCE OF DRINKING WATER WITH AVERAGE DISTANCE OF SOURCE : JULY—SEPTEMBER 1953

source of drinking water	government colonies		other localities		all localities		average distance of the source in yds.		
	number of households	percentage	number of households	percentage	number of households	percentage	government colonies	other localities	all localities
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. tube well	2	0.2	1	0.2	3	0.2	22.5	10.0	18.3
2. tap	729	82.9	394	84.1	1123	83.2	22.4	11.5	18.6
3. well	147	16.7	71	15.1	218	16.2	43.8	58.8	48.7
4. others	2	0.2	3	0.6	5	0.4	—	136.7	82.0
5. all types	880	100.0	469	100.0	1349	100.0	25.9	19.5	23.7

TABLE 34 : DISTRIBUTION OF HOUSEHOLDS BY SOURCE OF NON-DRINKING WATER WITH AVERAGE DISTANCE OF SOURCE : JULY—SEPTEMBER 1953

source of water (other than drinking water)	government colonies		other localities		all localities		average distance of the source in yds.		
	number of households	percentage	number of households	percentage	number of households	percentage	government colonies	other localities	all localities
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. tube well	2	0.2	1	0.2	3	0.2	22.5	10.0	18.3
2. tap	655	74.5	291	62.1	946	70.1	19.5	9.6	16.4
3. tank	2	0.2	11	2.3	13	1.0	125.0	293.6	267.7
4. well	127	14.5	50	10.7	177	13.1	42.9	55.7	46.5
5. others	10	1.1	9	1.9	19	1.4	32.0	224.4	123.2
6. not returned	84	9.5	107	22.8	191	14.2	—	—	—
7. all types	880	100.0	469	100.0	1349	100.0	21.4	23.1	22.0

TABLE 35 : DISTRIBUTION OF HOUSEHOLDS BY TYPE OF LATRINES USED WITH AVERAGE DISTANCE OF LATRINES : JULY—SEPTEMBER 1953

type of latrine available	government colonies		other localities		all localities		average distance of the latrines in yds.		
	number of households	percentage	number of households	percentage	number of households	percentage	government colonies	other localities	all localities
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. underground drainage system with flush	27	3.1	219	46.7	246	18.2	3.9	6.1	5.9
2. septic tank	149	16.9	6	1.3	155	11.5	26.9	29.2	27.0
3. old type	685	77.9	198	42.2	883	65.5	31.0	24.6	29.6
4. others	13	1.4	44	9.4	57	4.2	93.8	146.1	134.2
5. not returned	6	0.7	2	0.4	8	0.6	7.5		6.2
6. all types	880	100.0	469	100.0	1349	100.0	30.2	27.3	29.2

TABLE 36 : DISTRIBUTION OF HOUSEHOLDS BY NUMBER USING SAME SOURCE OF DRINKING WATER : JULY—SEPTEMBER 1953

number of households using same source	number of households			percentage		
	all sources	from tap only	from well only	all sources	from tap only	from well only
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			government colonies			
1	63	63	—	7.2	8.6	—
2	6	6	—	0.7	0.8	—
3	2	2	—	0.2	0.3	—
4	11	9	2	1.2	1.2	1.4
5—10	36	32	4	4.1	4.4	2.7
11—20	131	120	9	14.9	16.5	6.1
21—30	192	178	14	21.8	24.4	9.5
above 30	432	315	117	49.1	43.3	79.6
not returned	7	4	1	0.8	0.5	0.7
total	880	729	147	100.0	100.0	100.0
			other localities			
1	171	168	2	36.4	42.6	2.8
2	28	26	2	6.0	6.6	2.8
3	19	18	1	4.1	4.6	1.4
4	15	12	3	3.2	3.0	4.2
5—10	74	65	9	15.7	16.5	12.7
11—20	37	29	8	7.9	7.4	11.3
21—30	19	15	4	4.1	3.8	5.6
above 30	62	42	20	13.2	10.7	26.2
not returned	44	19	22	9.4	4.8	31.0
total	469	394	71	100.0	100.0	100.0
			all localities			
1	234	231	2	17.3	20.6	0.9
2	34	32	2	2.5	2.8	0.9
3	21	20	1	1.6	1.8	0.5
4	26	21	5	1.9	1.9	2.3
5—10	110	97	13	8.1	8.6	6.0
11—20	168	149	17	12.5	13.3	7.8
21—30	211	193	18	15.6	17.2	8.3
above 30	494	357	137	36.7	31.8	62.7
not returned	51	23	23	3.8	2.0	10.6
total	1349	1123	218	100.0	100.0	100.0

TABLE 37 : DISTRIBUTION OF HOUSEHOLDS BY NUMBER USING SAME SOURCE OF NON-DRINKING WATER : JULY—SEPTEMBER 1953

number of households using same source	number of households			percentage		
	all sources	from tap only	from well only	all sources	from tap only	from well only
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			government colonies			
1	36	36	—	6.4	8.5	—
2	6	6	—	0.7	0.9	—
3	1	1	—	0.1	0.2	—
4	10	10	—	1.1	1.5	—
5—10	23	21	2	2.6	3.2	1.6
11—20	94	87	5	10.7	13.3	3.9
21—30	166	155	11	18.9	23.7	8.7
above 30	353	245	108	40.1	37.4	85.0
not returned	171	74	1	19.4	11.3	0.8
total	880	655	127	100.0	100.0	100.0
			other localities			
1	125	121	3	26.7	41.7	6.0
2	21	20	1	4.5	6.9	2.0
3	17	16	1	3.6	5.5	2.0
4	15	14	1	3.2	4.8	2.0
5—10	50	43	6	10.7	14.7	12.0
11—20	28	22	6	6.0	7.6	12.0
21—30	17	11	6	3.6	3.8	12.0
above 30	42	27	15	9.0	9.2	30.0
not returned	154	17	11	32.7	5.8	22.0
total	469	291	50	100.0	100.0	100.0
			all localities			
1	181	177	3	13.4	18.7	1.7
2	27	26	1	2.0	2.7	0.6
3	18	17	1	1.3	1.8	0.6
4	25	24	1	1.9	2.5	0.6
5—10	73	64	8	5.3	6.8	4.5
11—20	122	109	11	9.0	11.5	6.2
21—30	183	166	17	13.6	17.5	9.6
above 30	395	272	123	29.4	28.9	69.4
not returned	325	91	12	24.1	9.6	6.8
total	1349	946	177	100.0	100.0	100.0

NATIONAL SAMPLE SURVEY : SURVEY OF DISPLACED PERSONS IN BOMBAY

TABLE 38 : DISTRIBUTION OF HOUSEHOLDS BY NUMBER USING SAME LATRINE : JULY—SEPTEMBER 1953

number of households using same latrine	number of households			percentage		
	all types	old type only	under-ground drainage with flush only	all types	old type only	under-ground drainage with flush only
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			government colonies			
1	38	18	10	4.3	2.6	37.1
2	32	15	8	3.6	2.2	29.6
3	5	2	2	0.6	0.3	7.4
4	90	26	2	10.2	3.8	7.4
5—10	35	28	3	4.0	4.1	11.1
11—20	93	75	1	10.6	10.9	3.7
21—30	145	134	—	16.5	19.6	—
above 30	426	385	1	48.4	56.2	3.7
not returned	16	2	—	1.8	0.3	—
total	880	685	27	100.0	100.0	100.0
			other localities			
1	116	28	86	24.8	14.1	39.3
2	51	24	26	10.9	12.1	11.9
3	32	17	13	6.8	8.6	5.9
4	30	15	13	6.4	7.6	5.9
5—10	98	42	52	20.9	21.2	23.8
11—20	49	23	23	10.4	11.6	10.4
21—30	20	16	3	4.3	8.1	1.4
above 30	24	20	1	5.1	10.1	0.5
not returned	49	13	2	10.4	6.6	0.9
total	469	198	219	100.0	100.0	100.0
			all localities			
1	154	46	96	11.4	5.2	39.0
2	83	39	34	6.2	4.4	13.8
3	37	19	15	2.7	2.2	6.1
4	120	41	15	8.9	4.6	6.1
5—10	133	70	55	9.8	7.9	22.4
11—20	142	98	24	10.5	11.1	9.8
21—30	165	150	3	12.2	17.0	1.2
above 30	450	405	2	33.5	45.9	0.8
not returned	65	15	2	4.8	1.7	0.8
total	1349	883	246	100.0	100.0	100.0

TABLE 39 : DISTRIBUTION OF HOUSEHOLDS BY PRESENT LIVELIHOOD CLASSES : JULY—SEPTEMBER 1953

present occupation in India	occupation preferred										
	cultivation, agriculture and allied subjects	crafts and cottage industry	big industry	trade	construction	transport	services	professional services	others	not returned	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
government colonies											
1. cultivation, agriculture and allied subjects	1	–	–	–	–	–	–	–	–	1	2
2. crafts and cottage industry	1	44	1	13	1	–	3	1	17	3	84
3. big industry	–	–	15	3	–	–	6	–	4	2	30
4. trade	2	2	–	199	2	–	20	1	74	3	303
5. construction	1	1	–	1	10	–	1	–	4	–	18
6. transport	–	1	–	3	1	17	5	–	8	–	35
7. services	4	7	4	20	–	1	92	–	46	8	182
8. professional services	–	–	–	4	–	–	1	29	14	1	49
9. others	–	6	–	16	2	2	18	1	95	15	155
10. not returned	–	2	–	4	1	–	2	–	9	4	22
11. total	9	63	20	263	17	20	148	32	271	37	880
other localities											
1. cultivation, agriculture and allied subjects	1	–	–	–	–	–	–	–	–	–	1
2. crafts and cottage industry	2	28	–	2	–	–	3	–	1	4	40
3. big industry	1	–	16	2	–	–	3	–	1	1	24
4. trade	–	2	2	153	–	1	7	–	20	13	198
5. construction	–	–	–	1	4	–	–	–	2	2	9
6. transport	–	1	–	4	–	16	–	–	–	3	24
7. services	1	–	–	8	–	–	56	1	11	9	86
8. professional services	1	–	–	1	–	–	2	16	5	2	27
9. others	–	1	–	7	–	1	7	1	27	12	56
10. not returned	–	–	–	1	–	–	1	–	1	1	4
11. total	6	32	18	179	4	18	79	18	68	47	469

TABLE 39 (*contd.*) : DISTRIBUTION OF HOUSEHOLDS BY PRESENT LIVELIHOOD CLASSES : JULY—SEPTEMBER 1953

present occupation in India	occupation preferred										
	cultivation, agriculture and allied subjects	crafts and cottage industry	big indust	trade	construction	transport	services	professional services	others	not returned	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
all localities											
1. cultivation, agriculture and allied subjects	2	–	–	–	–	–	–	–	–	1	3
2. crafts and cottage industry	3	72	1	15	1	–	6	1	18	7	124
3. big industry	1	–	31	5	–	–	9	–	5	3	54
4. trade	2	4	2	352	2	1	27	1	94	16	501
5. construction	1	1	–	2	14	–	1	–	6	2	27
6. transport	–	2	–	7	1	33	5	–	8	3	59
7. services	5	7	4	28	–	1	148	1	57	17	268
8. professional services	1	–	–	5	–	–	3	45	19	3	76
9. others	--	7	–	23	2	3	25	2	122	27	211
10. not returned	–	2	–	5	1	–	3	–	10	5	26
11. total	15	95	38	442	21	38	227	50	339	84	1349

TABLE 40 : DISTRIBUTION OF HOUSEHOLDS BY THE STATES PREFERRED FOR LIVING IN : JULY—SEPTEMBER 1953

State preferred	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. Uttar Pradesh	5	0.6	2	0.4	7	0.5
2. Bombay	581	66.1	389	83.0	970	71.9
3. Saurashtra	—	—	1	0.2	1	0.1
4. Madhya Pradesh	21	2.4	—	—	21	1.6
5. Madhya Bharat	3	0.3	—	—	3	0.2
6. Hyderabad	6	0.7	1	0.2	7	0.5
7. Bhopal	1	0.1	—	—	1	0.1
8. Rajasthan	3	0.3	—	—	3	0.2
9. Punjab	2	0.2	2	0.4	4	0.3
10. Delhi	8	0.9	7	1.5	15	1.1
11. no opinion	250	28.4	55	11.7	305	22.6
12. not recorded	—	—	12	2.6	12	0.9
13. total	880	100.0	469	100.0	1349	100.0

TABLE 41 : DISTRIBUTION OF HEADS OF HOUSEHOLDS BY THE EXTENT OF KNOWLEDGE OF LOCAL LANGUAGE : JULY—SEPTEMBER 1953

knowledge about local language*	government colonies		other localities		all localities	
	number of households	percentage	number of households	percentage	number of households	percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. can read, write and speak	46	5.2	83	17.7	129	9.6
2. can read and speak but cannot write	5	0.6	22	4.7	27	2.0
3. can speak but cannot read or write	99	11.3	80	17.1	179	13.3
4. understand but cannot read, write or speak	63	7.2	58	12.4	121	9.0
5. cannot read, write, speak or understand	655	74.3	224	47.7	879	65.1
6. not returned	12	1.4	2	0.4	14	1.0
7. total	880	100.0	469	100.0	1349	100.0

* Only local language i.e., Marathi or Gujrati but most of them have working knowledge of Hindustani.

NATIONAL SAMPLE SURVEY: SURVEY OF DISPLACED PERSONS IN BOMBAY

FACSIMILE OF THE SCHEDULE OF INVESTIGATION

SAMPLE SURVEY OF REFUGEE FAMILIES IN BOMBAY STATE: JUNE–SEPTEMBER 1953

CONDUCTED BY NATIONAL SAMPLE SURVEY, GOVT. OF INDIA

[I]

1 name of h. h. head..

2 name of division................................ 3 district............

4 name of census tract.........................

5 nature of tract—urban/rural. 6 h. h. lives in govt. colony—yes/no

7 postal address...

8 name of informant.................... 9 relation to head.....................

[IA]

1 sample unit 1st stage.........................

2 sample unit 2nd stage.........................

3 size of h. h.............4 no. of e. & e.d........

h. h. occupation | 5 at present...............................

h. h. occupation | 6 in Pakistan...............................

7 mother tongue..............................

[II] DEMOGRAPHIC AND OCCUPATIONAL PARTICULARS ON DATE OF SURVEY.........................

sr. no.	relation to head of household	year of migration from Pak.	sex: m=1; f=2	age last birthday in years	education standard (α)	economic status (β) e=1, e.d.=2, n.e.d.=3	employment status (γ)	principal means of livelihood	occupation code	industry code	of last month ending on ....during						net income in Rs. (00.00) in cash and/or kind from all occupations					
											1st week		2nd week		whole month		1st week		2nd week		whole month	
											in main occupation											
											no. of days employed	average no. of hrs. per day	no. of days employed	average no. of hrs. per day	no. of days employed	average no. of hrs. per day	from enterprise	from other activities	from enterprise	from other activities	from enterprise	from other activities
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
1	head																					
2																						
3																						
4																						
5																						
6																						
7																						
8																						
9																						
10																						

[III] TOTAL RECEIPTS IN LAST MONTH ENDING ON..................................

sr. no.	particulars of current receipts for household purposes only	amount Rs.
(1)	(2)	(3)
1	total income from occupations (from bl II).	
2	receipts from govt. for consumptions (doles)	
3	,, ,, other sources	
4	,, as loan (non-govt.)	
5	,, as help (remittance, gift)	
6	,, from home supply of enterprise produce	
7		
8	total	

[IV] OUTSTANDING LOAN AND REPAYMENT

sr. no.	outstanding loan as on date of survey....	amount Rs.
(1)	(2)	(3)
1	govt. loan	
2	private loan	
3	total	
	loan repaid during the year ending on....	
1	govt. loan	
2	private loan	
3	total	

[V] LAND AS ON DATE OF SURVEY.....................

sr. no.	land for	land possessed					
		total acre (00.00)	govt.		other		
			acre 00.00	annual rent Rs.	tenancy	acre (00.00)	annual rent Rs.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	cultivation						
2	house						
3	others						
4	tanks etc.						
5	total				×		

[VI] ENTERPRISE

sr. no.	item	agriculture	production of goods and services	trade	profession and as employee
(1)	(2)	(3)	(4)	(5)	(6)
	reference period ending on..................	year	quarter	quarter	quarter
1	specification				
2	cost of raw materials (Rs.)				
3	other costs (Rs.)				
4	total gross income (Rs.)				
5	balance (Rs.)				
6	consumed for h. h. purposes (cash or kind) (Rs.)				
7	share of h.h. in enterprise, percent				

(α) illiterate—1; merely literate—2; under matric—3; matric—4; graduate—5. (β) e=earner; e.d.=earning dependent; n.e.d.=non-earning dependent. (γ) employer—1; employee—2; own account—3; unpaid h.h. labour—4; old, infirm, child, women doing domestic work, student—5; pensioners, beggars—6; unemployed seeking employment for the first time—7; unemployed seeking employment (excl.7)—8; rentiers—9.

[VII] LOANS AND GRANTS (EXCLUDING DOLES) RECEIVED FROM GOVT UPTO DATE OF SURVEY............ ...

sr. no.	specification of loan and grant	loan-1, grant-2	loans and grants received				disbursement of the loan and grant received								loan repaid to date	interest repaid to date
			date of receipt of first instalment	no. of instalments in which total is received	date of receipt of last instalment	total receipt Rs.	for the purpose taken Rs.	for other prod. purposes Rs.	house hold consumption Rs.	given to others as			balance in hand if any Rs.		
										loan Rs.	help Rs.	repayment of loan (priv.) Rs.			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	,13)	(14)	(15)	;16)
1	homestead land*														
2	house building														
3	cultivation land*														
4	agriculture (incld. cattle & implement purchase recla., irrig., etc.)														
5	trade														
6	profession														
7	maintenance								×						
8	maintenance grant								×						
9															
10	total	×	×		×										

*if land received instead of cash, amount in acres (00.00)—1) cultivation.................. 2) homestead..................

[VIII] CONSUMPTION OF FOOD DURING LAST MONTH ENDING ON........................

sr. no.	items	consumption from				total consumption	
		home supply		other sources (purchase, barter etc.)			
		qty. std. srs. (00.00)	value Rs. (00.00)	qty. std. srs. (00.00)	value Rs. (00.00)	qty. std. srs. (00.00)	value Rs (00.00)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	wheat & prod.						
2	millet & prod.						
3	other cereals (incld. rice & prod.)						
4	total (items 1-3)						
5	pulses						
6	vegetables						
7	oils & dairy prod.						
8	salt						
9	all other food						
10	total (items 4–9)						

[IX] EXPENDITURE ON OTHER ITEMS DURING LAST MONTH ENDING ON..........

sr. no.	items	amount Rs (00.00)
(1)	(2)	(3)
1	clothing	
2	house rent (res.)	
3	house maintenance & tax (res.)	
4	fuel & light	
5	miscellaneous	
6	total (items 1–5)	
7	grand total (item 6 plus item 10 bl. VIII)	

[XI] HOUSING

sr. no.	items		
(1)	(2)		(3)
1	rented/owned		
2	no. of rooms		
3	fl. space (sq. ft.)		
4	construction or purchase value (Rs.)		
5	type of	plinth	
6		wall	
7		roof	
8		ighting	

[X] PRIVATE LOANS AS ON DATE OF SURVEY........ .....

loan no.	amount taken Rs.	year taken	amount now due Rs.	rate of interest	security, if any	purpose for taking
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1						
2						
3						
4						

[XII] WATER & LATRINE

sr. no.	items	type	distance yds.	no. of families using
(1)	(2)	(3)	(4)	(5)
1	drinking water			
2	other water			
3	latrine			

[XIII] HOUSEHOLD'S PREFERENCE

sr. no.	item	preference	reason
(1)	(2)	(3)	(4)
1	staying place		
2	occupation		

[XIV] REMARKS:

[XV]

1 name of investigator..................................

....

2 date.. .. ..

3 signature.......... ....

[XVI]

1 name of inspector..........................

2 date of scrutiny................... ....

3 signature..

MISCELLANEOUS

THE MULTIPLEX METHOD FOR LINEAR PROGRAMMING

By RAGNAR FRISCH
University Institute of Economics, Oslo

1. THE PROBLEM

In standard form the linear programming problem can be formulated as follows. Consider $(n+m)$ real variables

$$x_1, x_2, \ldots, x_{n+m} \qquad \ldots \quad (1.1)$$

satisfying m linearly independent linear equations.

The number of degrees of freedom is consequently n and it is always possible at least in one way to express all the variables in terms of a set of n basis variables, linearly independent amongst themselves. Let

$$x_u, x_v, \ldots, x_w \qquad \ldots \quad (1.2)$$

be such a basis set. The equations may then be written in the standard form

$$x_j = b_{j0} + \sum_{k\text{-}u,\, v \ldots w} b_{jk} x_k \qquad (j = 1, 2, \ldots, n+m) \qquad \ldots \quad (1.3)$$

where the b_{j0} and b_{jk} are constants. Obviously

$$b_{j0} = 0 \quad \text{and} \quad b_{jk} = 0 \begin{bmatrix} 1 & \text{if } k = j \\ 0 & \text{otherwise} \end{bmatrix} \quad \text{when } j = u,\ v \ldots w. \qquad \ldots \quad (1.4)$$

If the equations (1.3) are taken for all $j = 1, 2, \ldots, n+m$, we get a system of equations that are linearly dependent, but if we take (1.3) only for[1] $j = 1, 2 \ldots) u, v \ldots w (\ldots n+m$, we get a system of equations that are linearly independent. More precisely : If the coefficients b_{j0} and b_{jk} have any values whatsoever, m equations of the form (1.3) for $j = 1, 2 \ldots)\ u, v \ldots w (\ldots n+m$ are always linearly independent.

We consider a linear *preference function*

$$f = p_0 + p_u x_v + p_v x_v + \cdots + p_w x_w \qquad \ldots \quad (1.5)$$

where the p_k $(k = 0, u, v \ldots w)$ are any given constants positive, negative or zero.

It does not restrict generality if we assume that only basis variables occur in (1.5). Indeed, any linear function of all the $n+m$ variables will assume the form (1.5) when all the variables are expressed in terms of the basis variables by means of (1.3).

The linear programming problem is the problem of determining that one or those sets of values of the variables that will maximize (1.5) subject to two sets of conditions. In the first place the *equations* (1.3), and in the second place the non-negativity conditions expressed by the *inequalities*

$$x_j \geq 0 \qquad (j = 1, 2, \ldots, n+m). \qquad \ldots \quad (1.6)$$

A great variety of problems can be reduced to this standard form. If there should be one of the variables that is *not* subject to the non-negativity condition, this variable could be eliminated and the problem reduced to one with the same number of degrees of freedom but one equation and one variable less. A problem containing any linear inequality, that is the condition that a given linear function of the variables shall be non-negative, can be reduced to the above standard form simply by putting the linear function in question equal to a new variable that is entered in the list (1.1).

The inequalities (1.6) introduce *discontinuities* in the boundary conditions, and this makes the method of Lagrange multipliers, which serves so well in many maximum problems with side conditions, inapplicable here.

The neoclassical method for handling the linear problem is the simplex method due to George B. Dantzig. At the Oslo University Institute of Economics, considerable effort has been made to handle the problem in a different way in the hope of finding one or some methods that may be more advantageous in cases with a great number of variables, particularly in problems of the type occurring in macroeconomic planning.[1]

It is an easy form of mental exercise to imagine iteration processes of one form or another which may look nice on paper, but in most cases such methods do not converge, or if they do converge in principle, the convergency will in general be so slow as to make the method entirely useless in practice. At the Oslo Institute we have accumulated a considerable junk-pile of such methods. Some of the methods we have used, do, however, contain features that may make them useful under certain circumstances. Some aspects of methods which have actually been used successfully on medium sized examples, are described in a number of mimeographed memoranda from the Oslo Institute of which the following might be mentioned here.[2]

21 June 1954 : Methods of solving linear programming problems. Synopsis of a lecture to be given at the International Seminar on Input-Output Analysis, Varenna (Lake Como) June—July 1954.

18 October 1954 : Principles of linear programming. With particular reference to the double gradient form of the logarithmic potential method.

29 March 1955 : A labour saving method of performing freedom truncations in linear programming. Part I.

13 May 1955 : The logarithmic potential method of convex programming. With particular application to the dynamics of planning for national development. Synopsis of a communication to be presented at the international colloquium of econometrics in Paris 23—28 May 1955.

The present memorandum describes—without proofs—a method which we have recently applied quite successfully to numerical examples of medium size. It may be called the multiplex method as distinguished from the simplex method.

It should always be remembered that any mathematical method and particularly methods in linear programming must be judged with reference to the type of computing machinery available. In all our work we have been guided by the possibilities and desiderata in a situation where the available equipment consists of desk machines or IBM 602A (or its electronic improvement 626, or similar types of calculating punches) or electronic automatic computors with a small high speed memory, as for instance the Oslo machine NUSSE (whose name academically stands for "Norwegian Universal etc.", but is better understood when referred to the meaning of the Norwegian word nusse which can appropriately be translated as "small cute girl"). Our outlook may perhaps be changed when we get used to the super modern, high capacity electronic computor that will be available here from the middle of next year.

---

[1] In a national planning problem of some size, one may easily run into several hundred variables and perhaps a hundred or more degrees of freedom.

[2] Subsequently a great number of other memoranda have been produced. I also discussed the problem in my lectures at the Institute for Social Studies, The Hague, in the Spring of 1957.

2. THE INITIAL POINT

Start from a point in the interior of the admissible region, that is such a combination of the values of the basis variables (1.2) as will make all the $(n+m)$ variables (1.1) effectively positive (not zero). If necessary, use the $S(\lambda)$ method of Section 6 in "Principles of linear programming" for finding such a point.[1]

From the initial point move by some method in a direction which is influenced by the direction of the preference vector, i.e. the vector with components p_k $(k = u, v \dots w)$. A highly effective way of determining the direction of such a movement and at the same time taking account of the boundary conditions is the double gradient form of the logarithmic potential method, but this does involve a higher computational cost than simply to move, say, in the preference direction. At present I cannot say definitely which one of the alternatives for the initial step will be computationally most profitable when account is taken of what may happen in the subsequent steps. It is, of course, always tempting to start in the cheapest way.

The computations involved in a movement in the direction determined by the double gradient form of the logarithmic potential method, are described in detail in "Principles of linear programming" (in particular (12.29)—(12.31) and § 13). A movement in the preference direction is simply determined by putting

$$x_k = x_k^0 + \lambda p_k \qquad (k = u, v \dots w) \qquad \dots \quad (2.1)$$

x_k^0 being the coordinates of the initial point, and letting λ increase from 0 through positive values.

---

[1] The $S(\lambda)$ method is well adapted for mechanical computation. It is in essence a method of finding solutions to a set of linear inequalities. I suspect that from the procedures involved in the $S(\lambda)$ method we can also extract a necessary and sufficient criteria for the case where an admissible region with a positive content exists, but I have not had an opportunity to follow up this idea. One of my associates in a working team at the Indian Statistical Institute in the winter of 1954-55 suggested that when the linear programming problem is transformed into a problem of solving a set of linear inequalities (which is possible by a theorem of J. von Neumann), we may solve the whole linear programming problem by the $S(\lambda)$ method. I do not know how fruitful this idea is. I suspect it will depend on how large the number of degrees of freedom in the optimum point is. The pointset that in the original formulation was the optimum point set (possibly only a single point) will appear as the admissible region in the transformed problem. In order to apply the $S(\lambda)$ method to (1.6) we do not need to know beforehand the dimensionality of the pointset that actually satisfies (1.6).

The change in the other variables in the course of this movement is determined by

$$x_j = x_j^0 + \lambda d_j \qquad (j = 1, 2, \ldots, n+m) \qquad \ldots \quad (2.2)$$

where

$$d_j = \sum_{k=u,\, v \ldots w} b_{jk}\, p_k \qquad (j = 1, 2, \ldots, n+m). \qquad \ldots \quad (2.3)$$

The largest value permissible for λ when we shall stay within the admissible region is[1]

$$\lambda_{\max} = \operatorname*{Min}_{j} \lambda_j \;\Big|\; \lambda_j \geqq 0 \qquad \ldots \quad (2.4)$$

where

$$\lambda_j = \frac{x_j^0}{-d_j} \;\Big|\; d_j < 0. \qquad \ldots \quad (2.5)$$

The value (2.4) of λ determines the breaking-out-point where we must stop.

We consider also the values of λ_j that are positive larger than $\lambda_{\max}$, and *come closest to* $\lambda_{\max}$. This determines the priority order of what may be called the *optimum candidates* determined by the breaking-out-point considered. If $j = \alpha$ is the value of j that furnishes the minimum (2.4), $j = \beta$ the value that gives the next to smallest positive value of λ_j etc. ...up to some value $j = \gamma$, we say that the list $\alpha, \beta \ldots \gamma$ is the list of optimum candidates corresponding to the breaking-out-point considered. The variable x_α may be called the leading canditate. Some of the candidates may give the same value of λ. For instance it may happen that $\lambda_\alpha = \lambda_\beta$ (in the double gradient form this is always so) which expresses the fact that in the breaking-out-point we hit simultaneously two of the boundary planes.[2]

3. The number of candidates to consider

At each breaking-out-point which has been reached through a movement whose direction is to a large extent determined by (local) considerations on the maximization of the preference function, we pick a certain number of the optimum candidates and make provisionally the guess that there exists at least one optimum point where these candidates are zero. Much depends on a happy choice of this number. If we make it too *large*, we run a great risk of making a wrong guess and will then have to face the possibility that to correct the mistake we must make additional computations at the end of the work when it turns out that the point we have reached is not optimal (compare § 9 below). And if we make the number of variables included in the guess at each step too *small*, we shall proceed through many calculations that are in fact unnecessary. We can in principle never be completely protected against making a wrong guess. Even a leading candidate may finally turn out to be different from zero in the optimum point. But if we are willing to run *some* risk of having to make correctional computations in the end, we may very often make short cuts that can save us a good deal of work.

---

[1] An indication written after a vertical bar expresses the condition imposed on the suffix over which a maximization or minimization shall take place.

[2] The idea of using the values of the λ for ranging the candiates in a priority order was suggested some years ago by Mrs. Inger Haugstad, Chief computor in the University Institute of Economics, Oslo.

Since we have an exact necessary and sufficient criterion by which to decide whether a given point is optimal or not, and since we have a method of proceeding to correctional computations at the end if the optimality criterion should turn out in the negative, it is a sound procedure to work with guesses on the optimum candidates.

Empirically we have found that it pays fairly well to include at each round additional candidates equal in number to *the square root of the number of degrees of freedom with which we arrived in the breaking-out-point in question.*

Since this rule will usually give fractional numbers, we use a standard procedure for the rounding off to integers.

We first compute by the following recurrence formulae

$$N_t = N_{t-1} - \sqrt{N_{t-1}} \qquad \dots \quad (3.1)$$

the number of degrees of freedom N_t which should theoretically be retained in round number t, the initial value being $N_0 = n$.

These numbers N_t are then rounded off to the nearest integer

$$n_t = \text{nearest integer to } N_t \qquad \dots \quad (3.2)$$

and $n_{t-1} - n_t$ is taken as the number of new candidates to include in our optimality guess before we start on round number t.

Or otherwise expressed : When going from round number $t-1$ to round number t, we *reduce the number of degrees of freedom* by

$$n_{t-1} - n_t . \qquad \dots \quad (3.3)$$

For instance, if $n = 12$, we get the scheme indicated in Table (3.4).

TABLE (3.4). EXAMPLE OF COMPUTATIONS (3.1)—(3.3)

round number	solution of (3.1) N_t	number of degrees of freedom with which this round is to be performed n_t	number of variables to be put equal to zero when starting on this round $n_t - {}_{t-1}$
$t =$ 0	12.0000	12	0
1	8.5359	9	3
2	5.6143	6	3
3	3.2448	3	3
4	1.4435	1	2
5	0.2421	0	1

Towards the end of the work when the rule indicates that two variables shall be put equal to zero, we only take *one* (the leading candidate), because it will involve no more work to use (4.2) twice than to use first (5.2) (to bring one variable to zero) and then (4.2).

4. A preference increasing movement

Consider a point where $\nu(\leqslant n)$ variables

$$x_\alpha,\ x_\beta,\ \ldots,\ x_\gamma \qquad \ldots \quad (4.1)$$

—basis variables or dependent vriables—are equal to zero, all the other variables being effectively positive. That is to say we are in a point on the boundary of the admissible region.

Consider a movement from this point in the direction

$$d_k = p_k + B_\alpha\, b_{\alpha k} + B_\beta b_{\beta k} + \ldots + B_\gamma b_{\gamma k} \qquad (k = u,\ v \ldots w) \qquad \ldots \quad (4.2)$$

where the $B_s (s = \alpha, \beta \ldots \gamma)$—independent of k—are the solutions of the system of linear equations

$$M_{0s} + B_\alpha M_{\alpha s} + B_\beta M_{\beta s} + \ldots + B_\gamma M_{\gamma s} = 0 \qquad (s = \alpha,\ \beta \ldots \gamma) \qquad \ldots \quad (4.3)$$

where
$$M_{rs} = \sum_{k=u,\ v \ldots w} b_{rk} b_{sk} \qquad \begin{pmatrix} r = 0,\ \alpha,\ \beta \ldots \gamma \\ s = 0,\ \alpha,\ \beta \ldots \gamma \end{pmatrix}. \qquad \ldots \quad (4.4)$$

We assume that the boundary vectors $b_{\alpha k}, b_{\beta k} \ldots b_{\gamma k}$ are linearly independent so that (4.3) has a unique solution. The B are called the regression coefficients.

With the direction numbers (4.2) we perform a movement

$$x_k = x_k^0 + \lambda d_k \qquad (k = u,\ v \ldots w) \qquad \ldots \quad (4.5)$$

x_k^0 being the initial values and λ being a parameter that increases from 0 through positive values. During this movement all the variables change according to

$$x_j = x_j^0 + \lambda d_j \qquad (j = 1,\ 2,\ \ldots,\ n+m) \qquad \ldots \quad (4.6)$$

where
$$d_j = \sum_{k=u,\ v \ldots w} b_{jk} d_k \qquad (j = 1,\ 2,\ \ldots,\ n+m). \qquad \ldots \quad (4.7)$$

The preference function will also change. More precisely, for any positive value of λ it will have *increased* as compared to its initial value. We have indeed

$$f - f^0 = \lambda \sum_{k=u,\ v \ldots w} p_k d_k = \lambda \frac{\begin{vmatrix} M_{00} & M_{0\alpha} & M_{0\beta} & \ldots & M_{0\gamma} \\ M_{\alpha 0} & M_{\alpha\alpha} & M_{\alpha\beta} & \ldots & M_{\alpha\gamma} \\ M_{\beta 0} & M_{\beta\alpha} & M_{\beta\beta} & \ldots & M_{\beta\gamma} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ M_{\gamma 0} & M_{\gamma\alpha} & M_{\gamma\beta} & \ldots & M_{\gamma\gamma} \end{vmatrix}}{\begin{vmatrix} M_{\alpha\alpha} & M_{\alpha\beta} & \ldots & M_{\alpha\gamma} \\ M_{\beta\alpha} & M_{\beta\beta} & \ldots & M_{\beta\gamma} \\ \cdots & \cdots & \cdots & \cdots \\ M_{\gamma\alpha} & M_{\gamma\beta} & \ldots & M_{\gamma\gamma} \end{vmatrix}}. \qquad \ldots \quad (4.8)$$

By (4.4) the expression to the right in (4.8) is the ratio between two moment matrices and hence positive definite.

All the variables $x_\alpha, x_\beta, \ldots, x_\gamma$ on the contrary will remain *unchanged* during the λ variation because by (4.2) and (4.3)

$$x_s - x_s^0 = \lambda \sum_{k=u,\, v \ldots w} b_{sk} d_k$$

$$= \lambda(M_{0s} + B_\alpha M_{\alpha s} + B_\beta M_{\beta s} + \ldots + B_\gamma M_{\gamma s}) \quad (s = \alpha, \beta \ldots \gamma). \quad \ldots \quad (4.9)$$

The parenthesis to the right in (4.9) is zero by (4.3).

Since the rest of the variables are all effectively positive, we can let λ increase a certain non-zero amount before we hit any boundary plane. Its maximum value will be determined in exactly the same way as in (2.4)—(2.5) and we will consequently have a new determination of some basis candidates, their number being determined as explained in § 3.

This guess being made, we try to move to a new point where these new variables are zero as well as those that were previously brought to zero. In doing so we retain the *same basis form of the equations*, that is to say we retain the same basis variables (1.2), the reduction in the number of degrees of freedom being produced by conditions imposed on the direction of the movement. This is done as follows.

5. A VARIABLE-ANNIHILATING MOVEMENT

Suppose that we are in any point x_k^0 in the admissible region and want to move towards a *direction point* where the ν variables $x_\alpha, x_\beta, \ldots, x_\gamma$ have reached prescribed (positive, negative or zero) increments

$$\Delta x_\alpha, \Delta x_\beta, \ldots, \Delta x_\gamma \quad \ldots \quad (5.1)$$

the direction of the movement being further specified by the requirement that it shall have components of the form

$$d_k = C_\alpha b_{\alpha k} + C_\beta\, b_{\beta k} + \ldots + C_\gamma b_{\gamma k} \qquad (k = u, v \ldots w) \quad \ldots \quad (5.2)$$

where $C_\alpha, C_\beta, \ldots, C_\gamma$ are constants independent of k.

The values of these ν constants will be determined by the requirement that the ν increments (5.1) shall have prescribed values. Indeed the coefficients C must then be solutions of the system

$$C_\alpha M_{\alpha s} + C_\beta\, M \;\; + \ldots + C_\gamma M_{\gamma s} = \Delta x_s \qquad (s = \alpha, \beta, \ldots, \gamma). \quad \ldots \quad (5.3)$$

We assume that the matrix in the left member of (5.3) is non-singular so that the system has a unique solution.

Inserting the C values thus obtained into (5.2), we get a set of well defined direction numbers, and can proceed to a movement of the type indicated by (4.5)—(4.7). In particular we may choose the increments (5.1) as

$$\Delta x_s = -x_s^0 \qquad (s = \alpha, \beta, \ldots, \gamma) \quad \ldots \quad (5.4)$$

in which case all the ν variables $x_\alpha, x_\beta, \ldots, x_\gamma$ will be zero in the direction point.

As λ increases from 0 to 1, we pass from the initial point to the direction point where the prescribed increments (5.1) are reached. Two cases are now to be distinguished : We may be able to proceed unhindered up to the point $\lambda = 1$ where the increments (5.1) are realized, or we may en route be stopped by the boundary plane of a variable which is *not* in the set $x_\alpha, x_\beta, \ldots, x_\gamma$.

If the set $x_\alpha, x_\beta, \ldots, x_\gamma$ has been successfully selected, we will as a rule have the first alternative. In this case all is well, and we can proceed from the point thus obtained by using (4.2)—(4.3). Computationally this is now an extremely simple matter if the system (5.3) has been solved by the Gaussian elimination-algorithm.[1] The computation of the B_s $(s = \alpha, \beta \ldots \gamma)$ that satisfies (4.3) will then only be a matter of one more back solution.

In the other case we will meet (at least) one other boundary plane before reaching $\lambda = 1$. In this case we start from the point reached and use (4.2) with a set $x_\alpha, x_\beta, \ldots, x_\gamma$ that contains all the variables that are zero in this point. This is the variable which just stopped us together with those variables that remain zero during the movement up to the point where we now are. In the breaking-out-point to which this will lead us, a new selection of candidates can be made according to the standard rules of § 3 and § 4. Since the last movement was made by (4.2)—not by (5.2)—we will be justified in counting it as a round which means that a correspondingly larger number of variables will be included in the optimum group in the next round. In any case should the total net number of degrees of freedom n_t to be used in round number t be taken according to the rule of § 3.

6. THE ACCURACY OF THE COMPUTATIONS

As we proceed in the work the number of variables entering into the set $x_\alpha, x_\beta, \ldots, x_\gamma$ will increase. This means that the order of the system of linear equations—that is (4.3) or (5.3)—will increase. Since we do not need to consider an entirely new system at each round, but only have to add a certain number of rows and columns, the computational cost involved will not be prohibitive if the Gaussian algorithm is used. But some attention must be paid to the accuracy of the results. In general one will need all the more decimal-places the higher the system. One may find for instance that the variables that should in principle be maintained exactly equal to zero under the λ variation, do not arrive in the point $\lambda = 1$ with a value which is sufficiently close to zero. This is annoying for the subsequent work. The simplest way to handle this situation is to start by computing the moments (4.4) with great accuracy and in the subsequent work to make it a standard procedure at each step to improve the Gaussian algorithm values obtained for the B_s or C_s by at least one round of an *iteration process.*

Letting B_s and C_s be the values obtained through the Gaussian algorithm, the improved values B'_s and C'_s are then computed by

$$B'_s = \frac{-M_{os} - \sum\limits_{r<s} B'_r M_{rs} - \sum\limits_{r>s} B_r M_{rs}}{M_{ss}}. \qquad (s = \alpha, \beta \ldots \gamma) \qquad \cdots \quad (6.1)$$

$$C'_s = \frac{\Delta x_s - \sum\limits_{r<s} C'_r M_{rs} - \sum\limits_{r>s} C_r M_{rs}}{M_{ss}} \qquad (s = \alpha, \beta \ldots \gamma). \qquad \cdots \quad (6.2)$$

---

[1] And the elimination is performed in such a way as to store the triangular matrix obtained. This is always the case if the computations are done on desk machines, but it is not necessarily so when electronic automatic computors are used.

To perform one such iteration round is equivalent to testing the solutions by inserting in the original equations. The iteration is therefore also a reliable check. If judged necessary by the change produced from the uncorrected to the corrected values, one more iteration round may be performed. The computational cost involved in one iteration round is of the same order as that involved in the back solution by which B_s or C_s is obtained.

7. Sufficient criteria for optimality

In the work as here outlined each round will consist of two movements : First one of type (5.2) through which certain additional variables are brought down to zero, and next one of type (4.2) where all previous 0-variables are maintained as 0-variables and (usually) one more 0-variable is introduced, and at the same time a selection is made of certain variables which one will in the next step try to bring down to zero.

If the Gaussian elimination algorithm is used, each round will involve *adding* to the previous matrix as many rows and columns as there are variables which now are to be forced down to zero, elimination-treating these added rows and columns and then performing two back solutions of the *complete* system.

When (5.2) is used, the fact that the computation of the solution of (5.3) for the C_s $(s = \alpha, \beta \ldots \gamma)$ in the case where not all Δx_s are zero, proceeds regularly and does *not* by (5.2) give a set of d_k $(k = u, v \ldots w)$ that are all zero, assures us that the variables $x_\alpha, x_\beta, \ldots, x_\gamma$ are not linearly dependent.

When we reach a stage where all the d_k defined by (4.2) turn out to be zero for $k = u, v \ldots w$, we must consider the possibility that we have reached a point in the optimum region.

If all the coefficients $B_\alpha, B_\beta, \ldots, B_\gamma$ entering into the d_k in question are non-negative, we can say immediately that we have actually reached an optimum point and that there exists a $(n-\nu)$ dimensional region of such points, ν being the number of variables in the set $x_\alpha, x_\beta, \ldots, x_\gamma$.[1]

This is seen simply by noticing that if all the d_k defined by (4.2) are zero— a fact that only expresses a feature of the boundary vectors b_{sk} and the preference vector p_k and has nothing to do with the particular point, i.e. the particular values of $x_u, x_v, \ldots, x_w$, which we may happen to consider—we get by equating (4.2) to zero

$$f = p_0 + \sum_{k=u,\, v \ldots w} p_k x_k = p_0 - \sum_{s=\alpha,\, \beta \ldots \gamma} \; \sum_{k=u,\, v \ldots w} B_s b_{sk} x_k = p_0 - \sum_{s=\alpha\ \beta \ldots \gamma} B_s (x_s - b_{so})$$

that is

$$f = \left[p_0 + \sum_{s=\alpha,\, \beta \ldots \gamma} B_s b_{so} \right] - \sum_{s=\alpha,\, \beta \ldots \gamma} B_s x_s. \qquad \ldots \quad (7.1)$$

Since this formula holds good for any values of the basis variables inside or outside the admissible region or on its boundary, we can immediately conclude that if all the coefficients $B_s (s = \alpha, \beta \ldots \gamma)$ are non-negative, we can nowhere in the admissible region find a point which produces a value of f larger than the value obtained by putting

$$x_\alpha = x_\beta = \ldots = x_\gamma = 0. \qquad \ldots \quad (7.2)$$

---

[1] If N of the coefficients B are actually zero, there even exists an $(n-\nu+N)$ dimensional region of such points. The corresponding N conditions in (7.2) can then be dropped.

This optimum value for the preference function is

$$f_{opt} = p_0 + \sum_{s=\alpha,\,\beta\ldots\gamma} B_s b_{so}. \qquad \ldots \quad (7.3)$$

If the computation of the coefficients $B_\alpha, B_\beta, \ldots, B_\gamma$ has proceeded regularly, the boundary vectors $b_{\alpha k}, b_{\beta k}, \ldots, b_{\gamma k}$ cannot be linearly dependent. Hence the linear manifold defined by (7.2) must be of dimensionality $n-\nu$ where ν is the number of variables in (7.2). That is to say there exists (at least) one set of $n-\nu$ of the variables which are linearly independent and will, as they vary within the confines defined by the admissible region, generate a linear manifold where every point is optimal, i.e. produces the preference value (7.3).

If all the d_k defined by (4.2) are zero for $k = u, v \ldots w$, but one or more of the coefficients B are negative, we may have one or the other of the following two situations :

(I) We are actually in an optimal point, and if so, it must be possible to transform the expression (7.1) in such a way as to obtain a linear form with only non-negative coefficients B. This is the case where the optimum region is *multiply determined*, that is there exists at least one other set of ν variables such that the condition that *these* ν variables shall be zero, is equivalent to the condition (7.2).[1]

(II) We are not in an optimal point. One or more of our guesses of optimum candidates must then have been wrong, and in order to proceed towards the optimum point, we must again let loose one or more of the variables we have tentatively fixed at zero value.

Case (II) is discussed in § 9. In the present section we shall consider case (I). A few simple graphical illustrations will indicate the possiblities and what can be done in order to bring the expression (7.1) over into a form with only non-negative coefficients B so that a sufficient criterion for optimality emerges.

Fig. (7.4) illustrates a case where the admissible region is two-dimensional (being generated by the two basis variables x_u, x_v). The optimum region is the point $A(\nu = n = 2)$. The two straight lines that pass through A and *actually form the boundary* in the vicinity of A, are the two lines numbers (1) and (4), representing the conditions $x_1 = x_4 = 0$. There are, however, also two other lines passing through A, namely the lines numbers (2) and (3), representing the conditions $x_2 = 0$ and $x_3 = 0$. These two lines are situated entirely outside of the admissible region—with the exception of the point A—and the solution of the linear programming problem would have remained completely unchanged if one of or both the lines (2) and (3) had been omitted. They only add an unnecessary and undesirable complication in the picture. The optimum point can now be determined by putting *any two* of the four variables x_1, x_2, x_3, x_4 equal to zero. This might not be discovered until we get towards the end of the calculations and this is what causes the undesirable complication.

If the two lines numbers (2) and (3) had not been present, the situation in fig. (7.4) would have been very simple. The necessary and sufficient condition for the point A to be an optimum point, would then obviously have been that the preference vector p_k had been situated in the convex angle (the two-dimensional case of a convex polyhedral cone) between the two negative boundary vectors $(-b_{1k})$ and $(-b_{4k})$. They are denoted 1 and 4 and drawn

---

[1] When a region (in the special case $\nu=n$, a point) is multiply determined, it is customary to speak of "degeneracy". This term is not a happy one. It seems more appropriate and suggestive to speak of "a multiply determined" region.

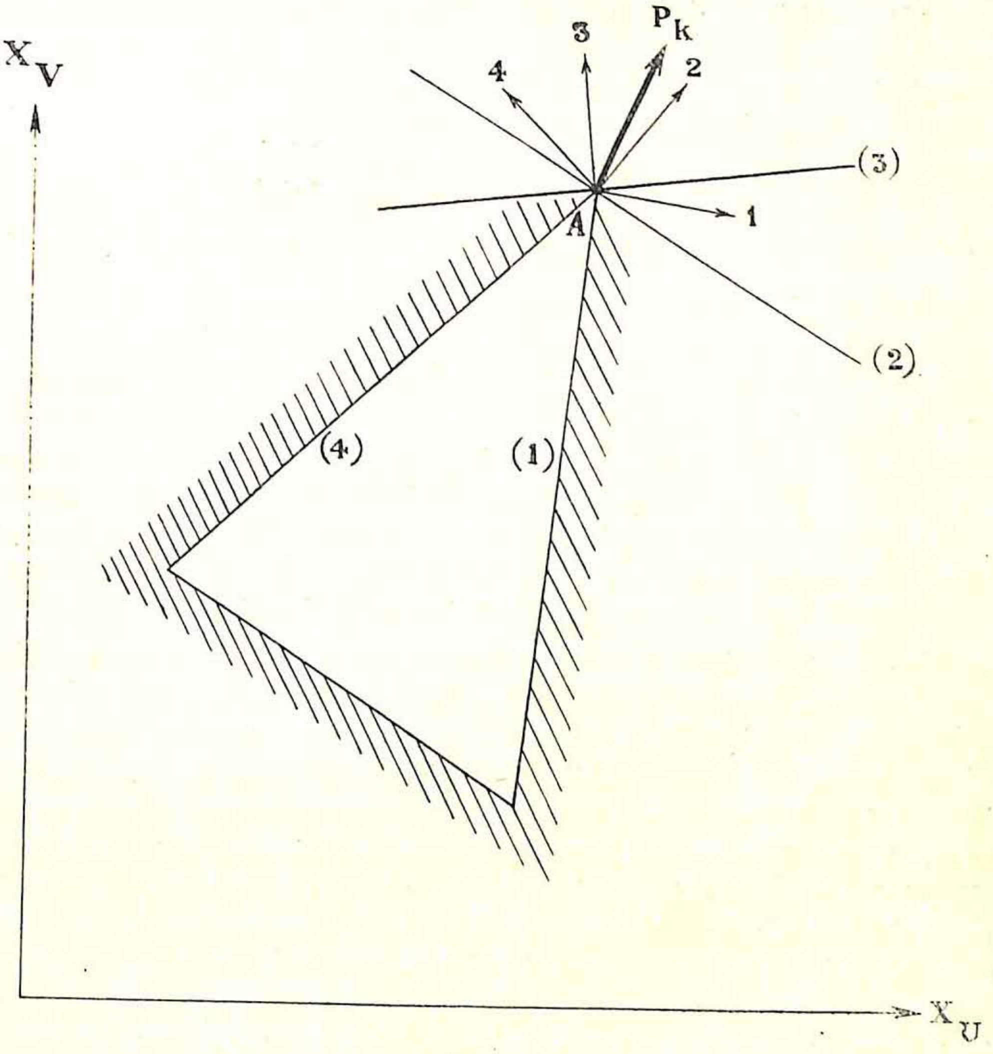


Fig. 7.4

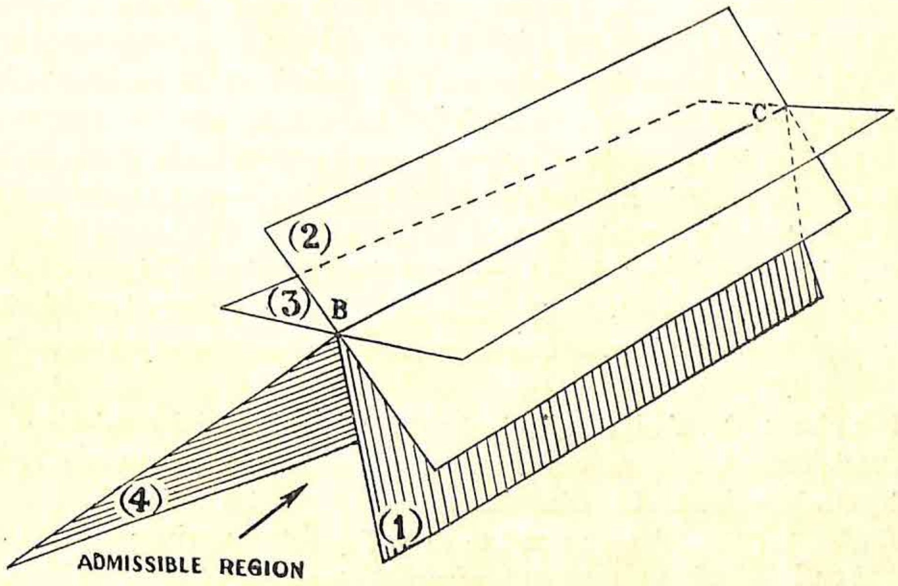


Fig. 7.5

as light arrows in fig. (7.4). This is the same as to say that the preference vector could have been expressed as a linear form in the two negative boundary vectors and with *non-negative* coefficients, i.e. as a linear form in the boundary vectors themselves, b_{1k} and b_{4k} and with *non-positive* coefficients. It is easily seen by a graphical inspection that this is the necessary and sufficient condition that there are no points that fall on the desired side of the preference plane through A and at the same time in the admissible region.

As the situation actually is in fig. (7.4) the preference vector may in $\binom{4}{2} = 6$ ways be expressed as a linear form in two boundary vectors $b_{\alpha k}, b_{\beta k} (\alpha \neq \beta)$, but not all these ways will give both coefficients of the boundary vectors non-positive. For instance, if we use the set 3, 4 the coefficient of b_{3k} will be negative but that of b_{4k} positive. If we use the set 2, 3 both coefficients will be negative, so that 2, 3 will actually furnish a *sufficient* criterion for optimality. Geometrically : If the preference vector p_k is situated in the angle 2, 3 it must *a fortiori* be situated in the angle 1, 4. If we use the set 1,4 we will also get a sufficient criterion. This latter set-being characterized by the fact that all the boundary vectors in A are included in the convex angle (1, 4) — has a further property not shared by any of the other sets, namely of furnishing a *necessary* criterion for optimality.

The above considerations show that if we are in an optimum point and this point is multiply determined, it will in general be possible to transform the expression for the preference *in different ways* to a linear function where all the coefficients of the variables are non–negative. Any of these transformations will give a sufficient criterion for optimality. The special set 1, 4 has the property of furnishing the *weakest possible* of the sufficient criteria for optimality. The preference vector in fig. (7.4) may indeed change to any direction between 1 and 4 without depriving the point A of its character of optimum point. The set of boundary vectors which has this property, we may call the minimum set. And the same designation may be used on the corresponding set of variables, that is x_1 and x_4.

Finally fig. (7.4) suggests that if we express the preference vector in terms of a set (θ, δ) which is not a minimum set and by so doing we get a positive coefficient of one of the vectors, say $b_{\theta k}$ and a negative coefficient for the other—in fig. (7.4) for instance $\theta = 4$, $\delta = 3$ —and we want to retain only minimum vectors, that is vectors permitting to express the optimum condition in its weakest form, it is not the variable x_θ whose vector $b_{\theta k}$ got a positive coefficient, that should be eliminated, but we should eliminate the variable which is such that the inclusion of its vector caused *some* coefficient or coefficients to become positive. For instance in fig. (7.4) if p_k is expressed in terms of b_{4k} and b_{3k} we will find that b_{4k} gets a positive and b_{3k} a negative coefficient. In this case it is x_3 that should be eliminated, not x_4, if we want to work towards a minimum set.

The situation is analogous if we have an optimum region of higher dimensionality. Fig. (7.5) illustrates the case $\nu = 2$, $n = 3$. There are four planes, any two of which define the optimum region, which is now one-dimensional, namely the segment of straight line BC.

The two planes (1) and (4) (shaded in fig. (7.5)) form the minimum set, while (2) and (3) only introduce unnecessary and undesired complications. Any two of the four planes will define the straight line that carries the segment BC which forms the optimal region. The vectors $b_{1k}, b_{2k}, b_{3k}, b_{4k}$ that are normal to these four planes, will—when translated to a common point on BC—be in a plane that is perpendicular to BC, and in this plane the situation is similar to the one exhibited around the point A in fig. (7.4).

This suggests the following heuristic rules for handling the case where we have reached a point which we suspect of being optimal and multiply determined.

Suppose that a movement is made by (4.2) and that the breaking out point to which this leads is determined by the fact that a certain number ω of the variables, namely

$$x_\phi \ \dots \ x_\psi \qquad (\phi \dots \psi \text{ being } \omega \text{ affixes}) \qquad \dots \quad (7.6)$$

become zero simultaneously.

Also suppose that a further movement of the form (4.2) but now with a subset

$$x_\theta \dots x_\delta \qquad (\theta \dots \delta \text{ being } \mu \text{ affixes amongst those in (7.6)}) \qquad \dots \quad (7.7)$$

included, i.e.,

$$d_k = p_k + B_\alpha b_{\alpha k} + \dots + B_\gamma b_{\gamma k} + B_\theta b_{\theta k} + \dots + B_\delta b_{\delta k} \qquad (k = u, v \dots w) \qquad \dots \quad (7.8)$$

makes all the d_k for $k = u, v \dots w$ equal to zero, but at least one of the B negative.

Regardless of the signs of the B, we know that if all the d_k are zero, the preference function can be written in the form

$$f = \Big[p_o + \sum_{s=\alpha,\, \beta \dots \gamma,\, \theta \dots \delta} B_s b_{so} \Big] - \sum_{s=\alpha,\, \beta \dots \gamma,\, \theta \dots \delta} B_s x_s. \qquad \dots \quad (7.9)$$

In the case now considered at least one B is negative.

We want to replace one or more of the variables numbers $\alpha, \beta \dots \gamma, \theta \dots \delta$ in (7.9) by an equal number of variables chosen among $\alpha, \beta \dots \gamma, \phi \dots \psi$ (compare (7.6)) in such a way as to obtain an expression for the preference function with only non-negative coefficients B.

In principle this can be done by testing all possible $\binom{\nu+\omega}{\omega-\mu}$ combinations of $(\nu+\mu)$ variables picked in the set consisting of the ν variables numbers $\alpha, \beta \dots \gamma$ and the variables (7.6). Since this may involve a tedious or even prohibitive work if done in a random order, we try to test the most promising combinations first.

For instance if one of the ν variables number $\alpha, \beta \dots \gamma$ is such that in the course of the previous work its inclusion produced one or more negative coefficients B, we may try to take this variable out. And then in a systematic way test one by one the $\binom{\omega}{\omega-\mu-1}$ alternatives obtained by adding $(\mu+1)$ of the variables (7.6).

In so doing we need to express some of the $(\nu+\omega)$ variables in terms of $(\nu+\mu)$ of them. In the multiply determined optimum region now considered, we can assume that any $(\nu+\mu+1)$ boundary vectors from the set of $(\nu+\omega)$ boundary vectors numbers $\alpha, \beta \dots \gamma$, $\phi \dots \psi$, are linearly dependent. For instance let numbers $\alpha, \beta \dots \gamma, \theta \dots \delta, \kappa$ be linearly dependent. That is we have

$$\sum_{s=\alpha,\, \beta \dots \gamma,\, \theta \dots \delta,\, \kappa} D_s b_{sk} = 0 \qquad \text{for all } k = u, v \dots w \qquad \dots \quad (7.10)$$

where the coefficients D are independent of k and not all zero.

They can be determined from the non-homogeneous system of order $(\nu+\mu)$

$$\sum_{s=\alpha,\, \beta \dots \gamma,\, \theta \dots \delta,\, \kappa} D_r M_{rs} = 0 \qquad (s = \alpha, \beta \dots \gamma, \theta \dots \delta) \qquad \dots \quad (7.11)$$

with the convention

$$D_\kappa = 1. \qquad \dots \quad (7.12)$$

If the matrix M_{rs} of rows and columns $\begin{pmatrix} r = \alpha, \beta \dots \gamma, \theta \dots \delta \\ s = \alpha, \beta \dots \gamma, \theta \dots \delta \end{pmatrix}$ is non-singular, (7.11) in conjunction with the convention (7.12) will determine the coefficients D uniquely. This assumption amounts to the same as to say that the $(\nu+\mu+1)$ set $(\alpha, \beta \dots \gamma, \theta \dots \delta, \kappa)$ is linearly dependent but the $(\nu+\mu)$ set $(\alpha, \beta \dots \gamma, \theta \dots \delta)$ is not.

If (7.10) holds, we have

$$\sum_{s=\alpha,\beta\dots\gamma,\theta\dots\delta,\kappa} D_s(x_s - b_{so}) = 0 \qquad \dots \quad (7.13)$$

Indeed, the left member of (7.13) is equal to

$$\sum_{k=u,v\dots w} x_k \sum_{s=\alpha,\beta\dots\gamma,\theta\dots\delta,\kappa} D_s b_{sk}$$

and the last sum in this expression is zero for all k by (7.10).

Through the regression equation (7.13)—which holds good for any point inside or outside of the admissible region or on its boundary—any of the variables that appear with a non-zero coefficient in (7.13) can be expressed in terms of the others. Choosing the affixes $\theta \dots \delta, \kappa$ differently, we obtain the necessary formulae for testing the various combinations and see if any of them can produce an expression of the form (7.9) with all the coefficients B non-negative. The first combination that yields this result, is sufficient to prove that the point considered is an optimum point.

A numerical example where the above method rapidly yields a sufficient criterion for optimality, is given in § 10.

If there is no combination that makes the sufficient optimality criterion verified, the region considered cannot be optimal.

If at any stage we find that by using (4.2) no boundary plane intervenes to stop the increase of λ to infinity, the linear programming problem is such that the preference function can be rendered arbitrarily great. In practice this is a trivial case.

8. CONTINUOUS WORK WITH NON-NEGATIVE REGRESSION COEFFICIENTS

In some cases the concrete setting of the problem may be such that it is not absolutely necessary to verify exactly that an optimum point has been reached. It may be sufficient to note that a *high* value of the preference function has been reached. If so, one does not need to worry about the trouble that may be involved in finding a sufficient criterion for optimality in the case where the optimal region is multiply determined. But if it is necessary to make definitely sure that no better solution exists, it may pay *already from the start* to be prepared for a very difficult case of a multiply determined optimal region, and to proceed accordingly. One way to do this is to try to work throughout only with non-negative regression coefficients.[1]

Suppose that one move of the kind (4.2) has been made and that all the coefficients B in (4.2) have—through the solution of (4.3)—turned out to be non-negative. And further suppose that when the move (4.2) with these B coefficient is performed, we reach a breaking point indicating that the additional set $\theta \dots \delta$ shall be included, so that the next move that contains the preference function, will be of the form (7.8). Finally, suppose that one or more of the coefficients B in (7.8) turn out to be negative. We may then decide *not* to make the move according to (7.8), but instead drop one or perhaps more of the variables numbers $\theta \ldots \delta$, and see if this should produce a determination of the direction numbers d_k which involve only non-negative coefficients B. We may first try to drop the variable that was the last candidate in the priority order for the variables numbers $\theta \ldots \delta$, and next—if necessary—try to drop that variable which was next in priority order, and so on. Finally—if necessary—we may try to drop more than one of the variables numbers $\theta \ldots \delta$. As soon as a set is found that gives only non-negative B, we proceed.

---

[1] Intuitively one would think that this is always possible because the admissible region is convex, but I have not gone through an exact proof.

In the new breaking out point thus determined, the number of new candidates to include in the optimum set may be determined *by starting a new count* on the basis of the square root rule formulated in the beginning of § 3.

By following the above rule in all the rounds of the work it will usually be easier to handle the optimality criterion at the end. This is illustrated in the example in § 10.

It is interesting to note how the value of the preference function is gradually increased. Towards the end the rate of increase gradually decreases from round to round in such a way that *most* of the increase can be realized without carrying the work through to the exact optimum. In many practical cases this may be of considerable value.

9. CORRECTIONAL COMPUTATIONS IN THE CASE WHERE THE OPTIMALITY CRITERION TURNS OUT IN THE NEGATIVE

If we have reached a *corner* on the boundary, that is with $\nu = n$ linearly independent variables $x_\alpha, x_\beta, \ldots, x_\gamma$ equal to zero, or we have reached a stage where $(n-\nu)$ is equal to the dimensionality of the optimum point, as we happen to know it beforehand, and if in either case the optimality criterion turns out in the negative, we must at one stage or another have made at least one wrong guess. At least one of the vriables must therefore be set free from the condition of being zero. The same applies in the case (II) of Section 7.

Sometimes the general aspect of the numerical results will give an indication of which one or which ones of the variables in the set $x_\alpha, x_\beta, \ldots, x_\gamma$ should be set free from the condition of being 0.

One way to set a variable free is, of course, to all make the computations over again with this variable kept out of the set $x_\alpha, x_\beta, \ldots, x_\gamma$. This may easily be done if the variable in question was one of the *latest* to be included in the set $x_\alpha, x_\beta, \ldots, x_\gamma$. Only a slight recomputation will then be needed.[1] But if the variable in question occurs somewhere in the first part of the list, the recomputation will virtually mean to start afresh. The problem therefore arises if it should be possible to perform the operation of setting one or more variables free by a computational procedure that permits us to use most of the previous results. This is done as follows.

---

[1] If the computations have been made in such a way as to store the triangular matrix which is obtained in the course of the work *or* the computations proceed by means of the inverse. This will often be the case when one works with an automatic electronic computor. If N is the order of the inverse which the machine can handle directly and the order of the matrix n is larger than N, perhaps very much larger, the best procedure is probably to increase stepwise the order of the matrix by means of formulae similar to those of § 4 in "Principles of linear programming". One will then begin by computing the inverse of the upper left $N \times N$ corner directly, end then compute the inverse of the $2N \times 2N$ upper left corner by means of one new $N \times N$ inversion and some extra work. Further the inverse of the upper left $3N \times 3N$ corner is computed by means of one new $N \times N$ inversion and some extra work. And so on. As one proceeds, the extra work needed becomes larger and larger in proportion to the $N \times N$ inversion, but even so this procedure is probably the most effective when the capacity of the machine is not sufficient to handle the whole inversion directly. The procedure is particularly adapted to the method of the present paper which proceeds in any case by adding rows and columns round by round. Inversion formulae are given in section 13.

Suppose that by an application of (5.2)—with not all the Δx_s in (5.3) equal to 0—we have reached a point where all the variables $x_\alpha, x_\beta, \dots, x_\gamma$ have actually been reduced to zero, and suppose that we subsequently decide that we should have wanted to make this movement *without making any assumptions* about a certain *subset* $x_\theta \dots x_\delta$ of the variables in the set $x_\alpha, x_\beta, \dots, x_\gamma$.

To do this we insert in the ν dimensional system (5.3) *not* the values $\Delta x_\theta \dots \Delta x_\delta$ that were prescribed in the movement as first defined, but such values $\Delta x_\theta \dots \Delta x_\delta$ as will make $C_\theta \dots C_\delta$ vanish in (5.2). These values of $\Delta x_\theta \dots \Delta x_\delta$ are determined by the system of equations

$$C'_r + C_r^\theta \, \Delta x_\theta + \dots + C_r^\delta \, \Delta x_\delta = 0 \qquad (r = \theta \dots \delta) \qquad \dots \quad (9.1)$$

where C'_s $(s = \alpha, \beta \dots \gamma)$ is the solution of (5.3) for $\Delta x_\theta = \dots = \Delta x_\delta = 0$ and all the other Δx_s inserted with the values they have in the movement now wanted made, C_s^θ $(s = \alpha, \beta \dots \gamma)$ the solution of (5.3) for $\Delta x_\theta = 1$ and all the other Δx_s equal to zero, and similarly for C_s^δ etc.

Having determined the solution $\Delta x_\theta \dots \Delta x_\delta$ of (9.1) the direction (5.2) with

$$C_s = C'_s + C_s^\theta \, \Delta x_\theta + \dots + C_s^\delta \, \Delta x_\delta \qquad (s = \alpha, \beta \dots \gamma) \qquad \dots \quad (9.2)$$

is the direction that would have been obtained by setting $x_\theta \dots x_\delta$ free.

Obviously if the C_s are determined by (9.2) the direction thus obtained will by (9.1) not contain the terms $C_\theta \, b_{\theta k} + \dots + C_\gamma b_{\gamma k}$.

Similarly, from a point where the variables $x_\alpha, x_\beta \dots x_\gamma$ are all zero, we may have made a movement by (4.2) but afterwards decide that we should have wanted to do it *without* including certain variables $x_\theta \dots x_\delta$ that form a subset within the set $x_\alpha, x_\beta, \dots, x_\gamma$.

To handle this case we first consider the problem of moving from a given point in a direction

$$d_k = A_0 p_k + A_\alpha b_{\alpha k} + A_\beta b_{\beta k} + \dots + A_\gamma b_{\gamma k} \qquad (k = 1, 2 \dots n) \qquad \dots \quad (9.3)$$

where the constants A are determined in such a way that the preference function f as well as the variables $x_\alpha, x_\beta, \dots, x_\gamma$ assume given positive, negative or zero increments

$$\Delta f, \ \Delta x_\alpha, \ \Delta x_\beta \dots \Delta x_\gamma. \qquad \dots \quad (9.4)$$

In order to achieve this it is obviously necessary and sufficient that $A_0, A_\alpha, A_\beta, \dots, A_\gamma$ form the solution of the $(\nu+1)$ order system

$$A_0 M_{0s} + A_\alpha M_{\alpha s} + A_\beta M_{\beta s} + \dots + A_\gamma M_{\gamma s} = \Delta x_s \quad (s = 0, \alpha, \beta, \dots, \gamma) \qquad \dots \quad (9.5)$$

where for simplicity we have denoted the preference function f by x_0.

If we are only interested in the *direction* of the movement, not in the absolute values of the components (9.3), it makes no difference whether all the coefficients A are multiplied by a common positive factor. We may for instance normalize the solution by putting Δf, that is Δx_0, equal to such a number as will make $A_0 = 1$. If we do this, and put

$$\Delta x_\alpha = 0, \qquad \Delta x_\beta = 0 \ldots \Delta x_\gamma = 0, \qquad \ldots \quad (9.6)$$

we get the coefficients B_α, $B_\beta \ldots B_\gamma$ that satisfy (4.3).

This being so, consider the case where we want to determine the direction that would have been obtained if we had *not* imposed the 0-condition on the particular subset $x_\theta \ldots x_\delta$. We do it simply by *not* putting $\Delta x_\theta \ldots \Delta x_\delta$ equal to zero but to *such* values as will make $A_\theta \ldots A_\delta$ vanish. In other words we consider the system (9.5) for

$$\left.\begin{array}{l} \Delta x_0 = \text{magnitude to be disposed of so as to make } A_0 = 1 \\ \Delta x_s = 0 \text{ for } s = \alpha, \beta \ldots)\theta \ldots \delta(\ldots\gamma \\ \Delta x_\theta \ldots \Delta x_\delta \text{ magnitudes to be disposed of so as to assure } A_\theta = \ldots = A_\delta = 0. \end{array}\right. \qquad \ldots \quad (9.7)$$

For a moment we may disregard the equation $s = 0$ in (9.5) and only consider the equations $s = \alpha, \beta \ldots \gamma$. These equations are completely specified by the data written in two last lines of (9.7). We may (if we take account of the condition $A_0 = 1$) write these ν equations

$$M_{0s} + A_\alpha M_{\alpha s} + A_\beta M_{\beta s} + \ldots + A_\gamma M_{\gamma s} = \Delta x_s \quad (s = \alpha, \beta \ldots \gamma) \qquad \ldots \quad (9.8)$$

where the Δx_s are given by (9.7).

Let, as before, $B_s (s = \alpha, \beta \ldots \gamma)$ be the solution obtained by putting *all* the Δx_s in (9.8) equal to zero—compare the previously considered system (4.3) and let A_s^θ $(s = \alpha, \beta \ldots \gamma)$ be the solution obtained by putting $\Delta x_\theta = 1$, but all the other $\Delta x_s = 0$, A_s^δ the solution obtained by putting $\Delta x_\delta = 1$ but all the other $\Delta x_s = 0 \ldots$ etc. The general solution of (9.8) can then be written

$$A_s = B_s + A_s^\theta \, \Delta x_\theta + \ldots + A_s^\delta \, \Delta x_\delta \quad (s = \alpha, \beta \ldots \gamma). \qquad \ldots \quad (9.9)$$

The conditions $A_\theta = \ldots = A_\delta = 0$ are consequently expressed by the system

$$B_s + A_s^\theta \, \Delta x_\theta + \ldots + A_s^\delta \, \Delta x_\delta = 0 \quad (s = \theta \ldots \delta). \qquad \ldots \quad (9.10)$$

The order of this system is not ν but equal to the number of variables we want to set free.

When $\Delta x_\theta \ldots \Delta x_\delta$ are determined from (9.10) the direction numbers to use will be

$$d_k = p_k + A_\alpha b_{\alpha k} + \ldots)\, A_\theta b_{\theta k} + \ldots + A_\delta b_{\delta k} (\ldots + A_\gamma b_{\gamma k} \quad (k = u, v \ldots w) \qquad \ldots \quad (9.11)$$

where the A_s $(s = \alpha, \beta \ldots)\theta \ldots \delta(\ldots \gamma)$ may be determined, either by inserting $\Delta x_\theta \ldots \Delta x_\delta$ into (9.9) or by performing one back solution of (9.8) with the values for $\Delta x_\theta \ldots \Delta x_\delta$ in the right member.

The computational cost involved in this procedure consists in : 1) As many back solutions of (9.8) as there are variables to be set free, and 2) one back solution to obtain the $A_s\ (s = \alpha, \beta ...)\ \theta\ ...\ \delta(...\gamma)$, or the corresponding work involved in using (9.9). The B_s can be assumed known from the previous work.

Once the back solutions needed to compute the $A_s\ (s = \alpha, \beta\ ...)\theta\ ...\ \delta(...\gamma)$ are available, we may work with the number of degrees of freedom thus established in the same way as we proceeded originally to find our way towards a set of optimum candidates.

If no specific variable or variables in the set $x_\alpha, x_\beta, ..., x_\gamma$ distinguish themselves as the ones that need to be set free, one may make a small move into the interior of the admissible region—even at the cost of some decrease in the preference function — and from the point thus obtained make a fresh start. If the move into the interior of the admissible region is made in reasonable way, one will get a new starting point from where it is possible to make a much more rapid progress than from the original starting point.

If $x_\alpha, x_\beta, ..., x_\gamma$ are the zero variables, all of which one now wants to set free, one may decide on a set of effectively positive increments (5.1) which are chosen in the light of the values that were encountered in the previous work. An application of (5.2) with the coefficients C determined by (5.3) will then lead to this direction point, or possibly to some new boundary plane that prevents us from reaching the chosen direction point. If so, we may decide to stop, say, half way before this new boundary plane is reached.

No matter how the new initial point is determined, we start from it in one of the ways described in § 2. When deciding upon optimum candidates, we will now most probably notice such a conspicuous distribution of the values of the λ_j defined by (2.5) as to make it possible to jump more or less directly to an optimum point. Perhaps the distribution of the λ_j will point out fairly clearly that particular or those few particular variables that should be let loose from their zero values and also point out some other that should be equal to zero. If so, we may revert to the point where we were before the movement into the interior of the admissible region took place, and from this point make a new movement utilizing (9.1)—(9.2), or a generalization of these formulae to the case where additional conditions are imposed in the form of prescribed increments for some further variables.

10. A NUMERICAL EXAMPLE

As an example we consider a problem in optimal blending of aviation gasolines. The problem is a small one, containing only 22 variables and 10 linearly independent equations, i.e. 12 degrees of freedom. The example is not given here in order to give a fair illustration of the amount of work involved in the multiplex method as compared to that involved in other methods. Such a comparison can only be made on really large examples. The purpose of the present example is only to illustrate the principles underlying the multiplex method. The computations are given in details so that anyone who is interested may check for himself how the computations proceed. The form of the intermediate results given, is that which appears when the work is done on desk machines. The tables can simply be looked upon as working sheets to be used for this kind of machines. When the work is done on more or less automatic computors, intermediate results are either stored on cards or tapes or not stored at all.

The concrete nature of the problem is described in an article in Econometrica[1] and will not be considered here. We simply take as our starting point the equations expressed in basis form. They are given in Table (10.1) on page number 20. The cells of that table give the coefficients b_{jk} of (1.3). All the 22 variables are subject to the condition of being non-negative. The problem is to maximize the linear preference function given in the bottom row of Table (10.1). The cells of this bottom row give the coefficients p_k of (1.5).

As a starting point in the interior of the admissible region we use in this example the *centre* of the admissible region, i.e. the point where the logarithmic potential is (approximately) maximum. This was done simply because the example has also been handled by the logarithmic potential method. We could just as well have determined an initial point by the $S(\lambda)$ method or simply by guesswork. The way in which the initial point is chosen is irrelevant for the application of the multiplex method and will not be further discussed here. The initial point from which we start in the present example is given in column 0 in Table (10.2).

We shall describe how the work is done in the case where we decide to work throughout with non-negative regression coefficients and also in the case where we work without any sign condition on the regression coefficients. We shall describe these two alternatives independently. This will lead to some repetitions in the tables, but the explanations can then be given in a clearer and simpler form.

Working throughout with non-negative regression coefficients : From the initial point given under 0 in Table (10.2) we move in the cheapest way, i.e. by putting $d_k = p_k$ and compute the other d_j by (2.3). These numbers are given in the first of the columns in the column sector 1. The magnitudes λ_j defined by (2.5) are listed in the next columns with the priority order indicated in the small column to the right of the figures for λ_j. The value of the various λ_j (for negative d_j) need only be computed with an accuracy sufficient to determine their ranking order, but $\lambda_{\max}$ as defined by (2.4) is computed with great accuracy and listed at the bottom of the λ_j column. By means of this value and (2.2) the values of the variables in the breaking out point are computed. They are listed in the last column under 1. By the rule illustrated in (3.4) we shall pick the first three candidates. They are—as will be seen from the λ_j column in the column § 1—the variables numbers 21, 20, 22, taken in this order.

We are therefore led to using (5.2) with the set $\alpha, \beta \ldots \gamma$ made up of numbers 21, 20, 22. This means that we must solve the system (5.3)— which now consists of three equations. The matrix of the coefficients are given in the upper left 3×3 corner of Table (10·3). Since the matrix is symmetric only a triangle is filled in. Usual sum checks for this 3×3 matrix are run and listed in a column in the right part of the table.

The solution of this 3×3 matrix by the Gaussian algorithm is given in the triangle above the principal diagonal in the upper left 3×3 matrix in Table (10.4). The numbers listed below the diagonal in this 3×3 matrix are the multipliers used in working out the solution.[2] The right member as given by (5.4) is listed in its original form as the first three elements in the column marked 2 in the right hand part of Table (10.3), and the corresponding elimination treated numbers are listed in the same place in Table (10.4). The result of the back solution in the form of C_{20}, C_{21}, C_{22} are listed on the line 2 in the lower left part of Table (10.4).

---

[1] W. W. Cooper and B. Mellon : Blending aviation Gasolines, *Econometrica*, April 1952.

[2] A detailed description of the work sheet for the Gaussian algorithm in the symmetric case is given in Section 17 of Principles of Linear Programming.

TABLE (10.1) BASIS EQUATIONS IN THE LINEAR PROGRAMMING PROBLEM FOR THE BLENDING OF AVIATION GASOLINE.

Variable No.	Constant term	Basis variable No.: 1	2	3	4	5	6	7	8	9	10	11	12	Sum check
1 =		1.0												1.0
2 =			1.0											1.0
3 =				1.0										1.0
4 =					1.0									1.0
5 =						1.0								1.0
6 =							1.0							1.0
7 =								1.0						1.0
8 =									1.0					1.0
9 =										1.0				1.0
10 =											1.0			1.0
11 =												1.0		1.0
12 =													1.0	1.0
13 =	3800.00	- 1.0	- 1.0	- 1.0										- 3.0
14 =	2652.00				- 1.0	- 1.0	- 1.0							- 3.0
15 =	4081.00							- 1.0	- 1.0	- 1.0				- 3.0
16 =	1300.00										- 1.0	- 1.0	- 1.0	- 3.0
17 =		14.0			3.0			- 6.0			15.0			26.0
18 =			16.5			2.0			- 4.0			17.0		31.5
19 =				7.5			- 7.0			-13.0			8.0	- 4.5
20 =		1.9			- 1.1			2.9			-13.6			- 9.9
21 =			1.9			- 1.1			2.9			-13.6		- 9.9
22 =				1.9			- 1.1			2.9			-13.6	- 9.9
Sum check	11833.00	15.9	18.4	9.4	1.9	0.9	- 8.1	- 3.1	- 1.1	-10.1	1.4	3.4	- 5.6	23.3
Preference function	f.	0.360092	0.888446	1.493446	0.360092	0.888446	1.493446	0.360092	0.888446	1.493446	0.360092	0.888446	1.493446	

TABLE (10.2). STEP BY STEP TABLE WHEN WORKING CONTINUOUSLY WITH NON-NEGATIVE REGRESSION COEFFICIENTS.

	0	1				2				3			
	x_J	d_J	λ_J		x_J	d_J	λ_J		x_J	d_J	λ_J		x_J
J = 1	112.8129	0.360092			123.2803	- 0.3457	356.6	7	122.9346	0.3942679			293.1977
2	140.3695	0.888446			166.1955	0			166.1955	0.9727674			586.2815
3	1072.2389	1.493446			1115.6516	- 8.0606	138.4	5	1107.5910	1.6351873			1813.7406
4	116.2668	0.360092			126.7342	0.2001			126.9343	0.3403059			273.8942
5	140.3808	0.888446			166.2068	0			166.2068	0.8396283			528.7972
6	170.6608	1.493446			214.0735	4.6667			218.7402	1.4113852			828.2417
7	114.5645	0.360092			125.0319	- 0.5277	236.9	6	124.5042	0.4122553			302.5351
8	149.4271	0.888446			169.2531	0			169.2531	1.0171472			608.5044
9	53.6579	1.493446			97.0706	- 12.3030	7.9	3	84.7676	1.7097881			823.1333
10	20.5147	0.360092			30.9821	2.4745			33.4566	0.1154642			83.3194
11	20.0400	0.888446			45.8660	0			45.8660	0.2848819			168.8712
12	54.0105	1.493446			97.4232	57.6970			155.1202	0.4788763			361.9212
13	2474.5787	- 2.741984	902.5	6	2394.8726	8.4063			2403.2789	- 3.0022226	800.5	2	1106.7801
14	2224.6916	- 2.741984	811.3	5	2144.9855	- 4.8668	440.7	8	2140.1187	- 2.5913194	825.9	3	1021.0669
15	3769.3505	- 2.741984	1374.7	8	3689.6444	12.8307			3702.4751	- 3.1391906	1179.4	4	2346.8271
16	1205.4348	- 2.741984	439.6	4	1125.7287	- 60.1715	18.7	4	1065.5572	- 0.8792224	1212.0	5	685.8682
17	1548.5145	9.362392			1820.6675	36.0442			1856.7117	5.7990995			4361.0316
18	2363.8300	27.986[illegible]0			3177.3503	0			3177.3503	18.5043222			11168.3740
19	6581.6975	- 6.720[illegible]7	979.4	7	6386.3406	528.3944			6914.7350	-16.0120266	431.8	1	0
20	139.6882	- 3.56[illegible]1	39.2	2	36.0606	- 36.0606	1.0	1	0	0			0
21	255.6778	- 8.795115	29.1	1	0	0			0	0			0
22	1270.5921	-14.785115	85.9	3	840.8068	-840.8068	1.0	2	0	0			0
Sum check	23993.0001	3.482293			24094.2258	-312.4288			23781.7970	8.2913951			27362.4054
Preference function f =	2542.7944				2908.9894				2972.3629				7739.9824
Notes:		In the initial step: $d_J = p_J$	λ_{max} = 29.068780				λ_{max} = 1.0000				λ_{max} = 431.846334		

TABLE (10.2). (Continued).

	4				5				6			
	d_J	λ_J		x_J	d_J	λ_J		x_J	d_J	λ_J		x_J
J = 1	377.4841			670.6818	- 0.852842	786.4	2	62.7978	- 62.7978	1.0	1	0
2	377.4841			963.7656	- 0.274343	3513.0	7	768.2212	- 73.2700	10.5	8	694.9512
3	351.8120			2165.5526	1.127185			2968.9810	136.0678			3105.0488
4	332.7853			606.6795	- 0.305514	1985.8	5	388.9170	361.1075			750.0245
5	332.7853			861.5825	0.193809			999.7245	- 87.9571	11.4	9	911.7674
6	355.4963			1183.7380	0.111705			1263.3585	- 273.1504	4.6	5	990.2081
7	- 5.2663	57.4	4	297.2688	0.437137			608.8488	- 130.4864	4.7	6	478.3624
8	- 5.2663	115.5	5	603.2381	1.042029			1345.9696	557.9044			1903.8740
9	27.7391			850.8724	0.784432			1409.9955	288.7681			1698.7636
10	24.6973			108.0167	- 0.001223			107.1448	- 65.8048	1.6	3	41.3400
11	24.6973			193.5885	0.168194			313.4731	115.8429			429.3160
12	26.3118			388.2330	0.315708			613.2615	102.6781			715.9396
13	-1106.7801	1.0	1	0	0			0	0			0
14	-1021.0669	1.0	2	0	0			0	0			0
15	- 17.2065	136.4	6	2329.6206	- 2.263598	1029.2	3	716.1861	- 716.1861	1.0	2	0
16	- 75.7064	9.1	3	610.1618	- 0.482679	1264.1	4	266.1206	- 152.7162	1.7	4	113.4044
17	6685.1914			11046.2230	-15.497504	712.8	1	0	0			0
18	7334.9783			18503.3523	- 5.447848	3396.5	6	14620.2661	-1647.1579	8.9	7	12973.1082
19	0			0	0			0	0			0
20	0			0	0			0	0			0
21	0			0	0			0	0			0
22	0			0	0			0	0			0
Sum check	14020.1698			41382.5752	-20.945352			26453.2661	-1647.1579			24806.1082
Preference function f =				9788.0899				12807.9275				13679.8885
Notes:		λ_{max} = 1.0000				λ_{max} = 712.774328				λ_{max} = 1.0000		

TABLE (10.2). (Continued).

	7				8				9				10
	d_J	λ_J		x_J	d_J	λ_J		x_J	d_J	λ_J		x_J	d_J
J = 1	0			0	0			0	0			0	0
2	- 0.57222052	1214.5	4	334.5088	- 0.53965451	619.9	2	4.8002	- 0.00526710			1.7821	0
3	0.57222062			3465.4912	0.53965460			3795.1998	0.00526725			3798.2179	0
4	- 0.37308526	2010.3	5	515.0177	- 0.50119147	1027.6	3	208.8056	- 0.36441190	573.0	1	0	0
5	0.11387714			983.4987	0.14912033			1074.6056	0.27564854			1232.5527	0
6	0.25920804			1153.4836	0.35207108			1368.5858	0.08876325			1419.4473	0
7	- 0.44929401	1064.7	2	195.3517	- 0.19010743	1027.6	4	79.2032	- 0.13822554	573.0	2	0	0
8	0.18818327			2022.4108	0.01672140			2032.6269	0.19430674			2143.9647	0
9	0.26111073			1863.2375	0.17338623			1969.1699	- 0.05608117			1937.0353	0
10	- 0.06562935	629.9	1	0	0			0	0			0	0
11	- 0.04902594	8756.9	6	398.4345	- 0.08388853	4749.5	6	347.1817	0.01840212			357.7261	0
12	0.11465529			788.1611	0.08388866			839.4139	- 0.01840195			828.8695	0
13	0			0	0			0	0			0	0
14	0			0	0			0	0			0	0
15	0			0	0			0	0			0	0
16	0			113.4044	0			113.4044	0			113.4044	0
17	0.59206803			372.9443	- 0.36292953	1027.6	5	151.2080	- 0.26388246	573.0	3	0	0
18	-10.80005836	1201.2	3	6170.1395	-10.09905022	611.0	1	0	0			0	0
19	0			0	0			0	0			0	0
20	0			0	0			0	0			0	0
21	0			0	0			0	0			0	0
22	0			0	0			0	0			0	0
Sum check	-10.20799032			18376.0838	-10.46197949			11984.2080	- 0.26388222			11833.0000	0
Preference function f =				14435.4771				15083.3391				15242.2877	
Notes:		λ_{max} = 629.901104				λ_{max} = 610.962354				λ_{max} = 573.0010			

Using these three C-values the d_k are computed by (5.2) and the other d_j by (4.7). The result is given in the first column under column section 2 in Table (10.2). Using (4.5) we find that we can unhindered reach the direction point where the three variables Nos. 20, 21, 22 are zero. The rest of the variables in this point ar obtained by puting $\lambda = 1$ in (4.6). The result is listed in the last column under column section 2.

TABLE (10.3). MATRIX DATUM TABLE (MOMENT TABLE) WHEN WORKING CONTINUOUSLY WITH NON-NEGATIVE REGRESSION COEFFICIENTS.

	Left members:												Sum checks for left members in move No.:						Right members:			
	s=20	21	22	19	13	14	1	15	(17)	10	18	4	2	4	6	8	9	10	0	Δ		
																				2	4	6
s=20	198.19	0	0	0	-1.90	1.10	1.90	-2.90	-198.10	-13.60	0	-1.10	198.19	197.39	- 1.71	182.79	182.79	181.69	3.5649108	- 36.0606	0	0
21		198.19	0	0	-1.90	1.10	0	-2.90	0	0	-213.65	0	198.19	197.39	194.49	194.49	-19.16	-19.16	8.7956154	0	0	0
22			198.19	-124.55	-1.90	1.10	0	-2.90	0	0	0	0	198.19	72.84	69.94	69.94	69.94	69.94	14.7851154	-840.8068	0	0
19				338.25	-7.50	7.00	0	13.00	0	0	0	0		213.20	226.20	226.20	226.20	226.20	6.7205070		0	0
13					3.00	0	-1.00	0	- 14.00	0	- 16.50	0		-10.20	-25.20	-11.20	-27.70	-27.70	2.7419840		0	0
14						3.00	0	0	- 3.00	0	- 2.00	-1.00		13.30	10.30	13.30	11.30	10.30	2.7419840		-1106.7801	0
1							1.00	0	14.00	0	0	0			15.90	1.90	1.90	1.90	- 0.3600920		-1021.0669	0
15								3.00	6.00	0	4.00	0			13.30	7.30	11.30	11.30	2.7419840			- 62.7978
(17)									466.00	▨	▨	▨			270.90	▨	▨	▨	(-9.3623920)			-716.1861
10										1.00	0	0				-12.60	-12.60	-12.60	- 0.3600920			0
18											561.25	0					353.10	353.10	-27.9860490			
4												1.00						- 1.10	- 0.3600920			
Sum check													594.57	683.92	774.12	672.12	797.07	793.87	3.6633836			

TABLE (10.4). MATRIX SOLUTION TABLE WHEN WORKING CONTINUOUSLY WITH NON-NEGATIVE REGRESSION COEFFICIENTS.

		1					Left members:						
		s=20	21	22	19	13	14	1	15	(17)	10	18	4
	s=20	198.19	0	0	0	-1.90	1.10	1.90	-2.90	-198.10	-13.60	0	-1.10
	21	0	198.19	0	0	-1.90	1.10	0	-2.90	0	0	-213.65	0
	22	0	0	198.19	-124.55	-1.90	1.10	0	-2.90	0	0	0	0
	19	0	0	0.62843736	259.97812681	-8.69403098	7.69128110	0	11.17753166	0	0	0	0
	13	0.00958676	0.00958676	0.00358676	0.03344139	2.65461499	0.28884344	-0.98178516	0.29038738	-15.89913716	- 0.13037994	-18.54821127	-0.01054544
	14	-0.00555023	-0.00555023	-0.00555023	- 0.02958434	-0.10880803	2.72271428	0.09628067	-0.31398937	- 0.17054564	0.08966951	1.20400097	-0.99274732
	1	-0.00958676	0	0	0	0.36984088	-0.03536202	0.61527619	0.14630203	10.02501712	0.07898921	- 6.90246268	0.04175085
	15	0.01463242	0.01463242	0.01463242	- 0.04299412	-0.10938964	0.11532219	-0.23778269	2.28936641	2.43707525	- 0.19318008	-4.68289979	-0.13935552
	(17)	0.99954589	0	0	0	5.98924408	0.06263809	-16.29352359	-1.06451953	6.81829629	▨	▨	▨
	10	0.06862102	0	0	0	0.04911444	-0.03293387	-0.12838009	0.08438146	▨	0.03095597	0.33065128	-0.06042504
	18	0	1.07800595	0	0	6.98715683	-0.44220614	11.21847845	-2.04550035	▨	-10.68134127	130.25654995	1.76416957
	4	0.00555023	0	0	0	0.00397249	0.36461678	-0.06785709	0.06087078	▨	1.95196726	- 0.01354381	0.47872548
B in move No.:	3	0.01798734	0.04437971	0.07460071									
	5	0.02656730	0.05295968	0.13660615	0.08501321	1.26341199	0.63638197						
	7a	-0.06734733	0.07043042	0.13359925	0.05242820	1.58137988	0.65471006	2.55494227	0.99260273	- 0.08683532			
	7b	0.03130304	0.06893176	0.13399507	0.05544278	1.59163685	0.69874392	1.17206906	0.90016483	▨			
	8	0.18236592	0.07149519	0.13185051	0.04795120	1.56994137	0.66068096	0.85735411	1.07906065	▨	2.12008443		
	9	0.12147286	0.16088879	0.16435892	0.09042917	2.47867947	0.52088375	1.68778900	0.85058883	▨	1.29193889	0.07753198	
	10	0.29882509	0.14935468	0.14935469	0.06722224	2.28138660	0.85860020	1.46752691	1.05268523	▨	2.88792912	0.06722222	0.76121572
C in move No.:	2	-0.181950	0	-4.242428									
	4	-1.815979	-1.815979	-3.671656	-2.952842	-380.934473	-330.787773						
	6	291.108111	-8.517857	6.613869	24.078337	57.086056	97.326728	-1192.531659	-582.606293	259.551037			

TABLE (10.4) (continued).

	sum checks for left members in move no.:						right members :			
								Δ		
	2	4	6	8	9	10	0	2	4	6
$s=20$	198.19	197.39	− 1.71	182.79	182.79	181.69	3.56491080	− 36.0606	0	0
21	198.19	197.39	194.49	194.49	−19.16	−19.16	8.79561549	0	0	0
22	198.19	72.84	69.94	69.94	69.94	69.94	14.78511540	−840.8068	0	0
19		258.97537730	270.15290896	270.15290896	270.15290896	270.15290896	16.01202589		0	0
13		2.94345730	−13.64707762	2.12167960	−16.42659168	−16.43707711	3.53768715		−1106.7801	0
14		2.72271404	2.33445969	2.59467485	3.79867582	2.80592850	1.73268546		− 900.6405	0
1			10.78659495	0.84056705	− 6.06189564	− 6.02014479	0.85284213			− 62.797800
15			4.72644114	2.09618581	6.77908560	6.63973008	2.06080713			−701.253870
(17)			6.81829506				(−0.59206893)			1769.69587
10				0.03095581	0.36160709	0.30118205	0.06562927			
18					130.25653890	132.02070854	10.09904824			
4						0.47872294	0.36441184			

From the point thus obtained we perform a preference increasing movement, holding constant the three variables which have now reached zero, namely numbers. 20, 21, 22. To do this we have to use the *same* 3×3 matrix which we just used, but now with the right member given as the first three elements in the column marked 0 in the right hand part of Table (10.3) and Table (10.4).[1] The result is listed as the three coefficients B on line 3 in the lower left part of Table (10.4). Since all these three regression coefficients B turn out to be non-negative, we proceed. The coefficients d_j now to be used are listed in the first column in column section 3 in Table (10.2) and the λ_j listed in the subsequent column. The breaking out point defines a certain priority order of the remaining candidates. By (3.4) we shall also now pick three such candidates. They turn out to be numbers 19, 13, 14, taken in this order. We are thus led to consider the 6×6 system in the upper left corner of Table (10.3) and Table (10.4) with a right member which have the first four items equal to zero and the following two equal to the values listed on the fifth and sixth place in the right member column marked 4 of Table (10.3). The corresponding elimination treated elements are listed in the same places in Table (10.4). The back solution yields the six coefficients C listed on line 4 in the lower left part of Table (10.4). Computing the d_j — listed in the first column under column section 4 in Table (10.2)—we find that we can unhindered reach the direction point where all six variables numbers 20, 21, 22, 19, 13, 14 are zero. The values of all the variables in this point are listed in the last column in column section 4 in Table (10.2).

From the point thus obtained we want to perform a preference increasing movement holding constant all the six variables that have so far reached zero, namely numbers 20, 21, 22, 19, 13, 14. To do this we have to use the *same* 6×6 matrix as we just used, but now with the right member given as the first six elements in the column marked 0 in the right hand part of Tables (10.3) and (10.4). The result is listed as the six coefficients B on line 5 in the lower left part of Table (10.4). Since all these six regression coefficients B turn out to be non-negative, we proceed. The breaking out point is determined through the d_j built on these six B and through the corresponding λ_j. The values of the x_j in this breaking out point are listed in the last column in the column section 5 of Table (10.2). We select againt hree new candidates, namely numbers 17, 1, 15, of which the last two have to be forced down to zero. This is done by a 9×9 system with the right member indicated in column 6 in the right hand part of Tables (10.3) and (10.4). The resulting nine C coefficients are given on line 6 in the lower left part of Table (10.4). Using (4.5) we find that we can unhindered reach the direction point where all the nine variables numbers 20, 21, 22, 19, 13, 14, 17, 1, 15 are zero. The values of all the variables in this point are listed in the last column in column section 6 in Tab. (10.2).

From the point thus obtained we attempt again a preference increasing movement. This time, however, we find that *not* all the corresponding regression coefficients — namely the nine coefficients B listed on line 7a in the lower left part of Table (10.4)—are non-negative. We therefore decide that the preference increasing movement shall *not* be carried through with all the above nine variables equal to zero. The first candidate which came into the picture and produced some negative numbers in the list of regression coefficients was the leading candidate in the last breaking out point, namely number 17. This variable was also the only one of the newly added variables that got itself a negative regression coefficient. We therefore decide to throw number 17 out of the set of nine zero variables and perform the preferenc increasing movement by keeping only the remaining eight variables equal to zero, namely numbers 20, 21, 22, 19, 13, 14, 1, 15.

---

[1] If the work is done on desk machines and the tables actually used as work sheets, the paper is folded so that the right member column to be used always comes in convenient proximity to the other columns used.

To do this we must use an 8×8 system. It so happens that number 17 was added as the last row and column in the Tables (10.3) and (10.4). To carry the computations through without number 17 is therefore an easy matter. The result is the eight regression coefficients listed on line 7b in the lower left part of Table (10.4). Since all these eight coefficients turn out to be non-negative we proceed.

To decide on the number of degrees of freedom to retain in the next round we note that in the last breaking out point—the one built on the B coefficients on line 7b—we arrived with 12—8 = 4 degrees of freedom. According to the square root rule we should consequently in the next move put 2 variables equal to zero. As it involves just as much work first to force one variable to zero (the leading candidate is already zero) and next to make a preference increasing move—as to make two separate preference increasing moves, the latter alternative is chosen. This leads through the moves whose column sections in Table (10.2) are designated 7, 8 and 9 to the final attempt number 10 which produces a situation where all the d_k are zero, see the column number 10 in Table (10.2). Since the eleven coefficients B that lead to these zero d_k are all non-negative —as will be seen from line 10 in the lower left part of Table (10.4)—we can conclude that an optimal point has been reached.

Since exactly eleven B coefficients in this last round are effectively positive, we know that the optimal region is 12—11 = 1 dimensional. This one dimensional region may be generated in different ways. In Table (10.5) we have used the variable No. 6 to generate the

TABLE 10.5. OPTIMAL SOLUTION OF THE PROBLEM OF BLENDING AVIATION GASOLINES. ONE DEGREE OF FREEDOM GENERATED BY x_6, WHERE x_6 IS LYING IN THE INTERVAL (10.6)

		constant term	x_6	values of the variables in the lower and upper endpoints of the interval (10.6)	
				(1)	(2)
variable no.	$j=1$	0		0	0
	2	−65.3310	0.047281	0	60.0582
	3	3865.3310	—0.047281	3800.0000	3739.9418
	4	0	0	0	0
	5	2652.0000	—1.000000	1270.2399	0
	6	0	1.000000	1381.7601	2652.0000
	7	0	0	0	0
	8	1131.6719	0.713160	2117.0879	3022.9722
	9	2949.3281	−0.713160	1963.9121	1058.0278
	10	0	0	0	0
	11	17.6853	0.239559	348.6984	652.9958
	12	1168.9103	−0.239559	837.8972	533.5998
	13	0	0	0	0
	14	0	0	0	0
	15	0	0	0	0
	16	113.4044	0	113.4044	113.4044
	17	0	0	0	0
	18	0	0	0	0
	19	0	0	0	0
	20	0	0	0	0
	21	0	0	0	0
	22	0	0	0	0

optimum region. It is easy to see that when the eleven variables specified in the last move shall be zero, the point will be situated in the admissible region when and only when x_6 lies between the limits

$$1381.76 \leqslant x_6 \leqslant 2652.00 \qquad \dots \quad (10.6)$$

In other words we can attribute to x_6 any value between the limits (10.6) and then by Table (10.5) compute the corresponding values of all the variables. Any such point will be situated in the admissible region and will produce the same value of the preference function, namely $f = 15242.29$ and in all other admissible points than those generated in this way, will the preference function be actually less than the value indicated.

In the present example we did not need to use the procedure suggested in § 9.

In parameter form the solution (10.5)—(10.6) can be expressed by considering the two *endpoints* of the interval (10.6). In the lower endpoint, denoted (1), the variable x_2 is zero and in the upper endpoint, denoted (2), the variable x_5 is zero. The complete list of the values of all the variables in these two points are given in columns (1) and (2) in Table (10.5). Any point on the straight line segment generated by the variation of x_6 between the two limits considered can be expressed by introducing two non-negative parameters t_1 and t_2 subject to the condition

$$t_1+t_2 = 1. \qquad \dots \quad (10.7)$$

If $x^{(1)}_j$ and $x^{(2)}_j$ are the values of the variable x_j $(j = 1, 2, \dots n+m)$ in the two endpoints of the segment, i.e. the values read off in columns (1) and (2) respectively in Table (10.5), any point on the optimal segment can be expressed in the form

$$x_j = t_1\, x^{(1)}_j+t_2 x^{(2)}_j \qquad (j = 1, 2 \dots n+m). \qquad \dots \quad (10.8)$$

That is to say if we choose any two non-negative numbers t_1 and t_2 satisfying (10.7) and insert them in (10.8) we get a point on the optimal segment. Conversely any point on the optimal segment can be expressed in the form (10.8) where t_1 and t_2 are two non-negative numbers satisfying (10.7).

This parameter form has certain advantages for some practical purposes, for instance if one wants to add certain *secondary* preference considerations for comparing alternative points within the optimum region. If the optimum region is of dimensionality δ a similar expression with $(\delta+1)$ non-negative parameters adding up to unity, applies.

TABLE (10.9). STEP BY STEP TABLE WHEN WORKING WITHOUT SIGN CONDITIONS ON THE REGRESSION COEFFICIENTS.

	6	11				12				13
	x_j	d_j	λ_j		x_j	d_j	λ_j		x_j	d_j
j = 1	0	0			0	0			0	0
2	694.9512	- 0.559116	1242.9	2	3.6957	-0,00419965	880.0	4	1.7822	0
3	3105.0488	0.559116			3796.3043	0,00419976			3798.2178	0
4	750.0245	- 0.481042	1559.2	3	155.2949	-0,34083860	455.6	1	0	0
5	911.7674	0.156263			1104.9604	0.28003680			1232.5525	0
6	990.2081	0.324779			1391.7447	0.06080173			1419.4475	0
7	478.3624	- 0.306806	1559.2	4	99.0469	-0.21738513	455.6	2	0	0
8	1903.8740	0.100091			2027.6207	0.25534842			2143.9647	0
9	1698.7636	0.206715			1954.3324	-0.03796329			1937.0353	0
10	41.3400	- 0.026514	1559.2	5	8.5596	-0.01878633	455.6	3	0	0
11	429.3160	- 0.069408	6185.4	6	343.5047	0.03121257			357.7260	0
12	715.9396	0.095922			834.5313	-0.01242615			828.8696	0
13	0	0			0	0			0	0
14	0	0			0	0			0	0
15	0	0			0	0			0	0
16	113.4044	0			113.4044	0			113.4044	0
17	0	0			0	0			0	0
18	12973.1082	-10.493187	1236.3	1	0	0			0	0
19	0	0			0	0			0	0
20	0	0			0	0			0	0
21	0	0			0	0			0	0
22	0	0			0	0			0	0
Sum check	24806.1082	-10.493187			11833.0000	0			11833.0000	0
Preference function f =	13679.8885				15099.3550				15242.2878	
Notes:			λ_{max} = 1236.33632				λ_{max} = 455.6259			

TABLE (10.10). MATRIX DATUM TABLE (MOMENT TABLE) WHEN WORKING WITHOUT SIGN CONDITIONS ON THE REGRESSION COEFFICIENTS.

	Left members:												Sum checks for left members in move No.:						Right members:			
																				△		
	s=20	21	22	19	13	14	1	15	17	18	4	7	2	4	6	12	13	14	0	2	4	6
s=20	198.19	0	0	0	-1.90	1.10	1.90	-2.90	-198.10	0	-1.10	2.90	198.19	197.39	- 1.71	- 1.71	- 2.81	0.09	3.5649108	- 36.0606	0	0
21		198.19	0	0	-1.90	1.10	0	-2.90	0	-213.65	0	0	198.19	197.39	194.49	-19.16	-19.16	-19.16	8.7956154	0	0	0
22			198.19	-124.55	-1.90	1.10	0	-2.90	0	0	0	0	198.19	72.84	69.94	69.94	69.94	69.94	14.7851154	-840.8068	0	0
19				338.25	-7.50	7.00	0	13.00	0	0	0	0		213.20	226.20	226.20	226.20	226.20	6.7205070		0	0
13					3.00	0	-1.00	0	- 14.00	- 16.50	0	0		-10.20	-25.20	-41.70	-41.70	-41.70	2.7419840		-1106.7801	0
14						3.00	0	0	- 3.00	- 2.00	-1.00	0		13.30	10.30	8.30	7.30	7.30	2.7419840		-1021.0669	0
1							1.00	0	14.00	0	0	0			15.90	15.90	15.90	15.90	- 0.3600920			- 62.7978
15								3.00	6.00	4.00	0	-1.00			13.30	17.30	17.30	16.30	2.7419840			-716.1861
17									466.00	0	3.00	-6.00			270.90	270.90	273.90	267.90	- 9.3623920			0
18										581.25	0	0				353.10	353.10	353.10	-27.9860490			
4											1.00	0					1.90	1.90	- 0.3600920			
Sum check													594.57	683.92	774.12	899.07	901.87	897.77	4.0234756			

TABLE (10.11). MATRIX SOLUTION TABLE WHEN WORKING WITHOUT SIGN CONDITIONS ON THE REGRESSION COEFFICIENTS.

		Left members:											
		s=20	21	22	19	13	14	1	15	17	18	4	7
	s=20	198.19	0	0	0	- 1.90	1.10	1.90	- 2.90	-198.10	0	-1.10	2.90
	21	0	198.19	0	0	- 1.90	1.10	0	- 2.90	0	-213.65	0	0
	22	0	0	198.19	-124.55	- 1.90	1.10	0	- 2.90	0	0	0	0
	19	0	0	0.62843736	259.97812681	-8.69403098	7.69128110	0	11.17753166	0	0	0	0
	13	0.00958676	0.00958676	0.00958676	0.03344139	2.65461499	0.28884344	-0.98178516	0.29038738	-15.89913716	-18.54821127	-0.01054544	0.02780160
	14	-0.00555023	-0.00555023	-0.00555023	-0.02956434	-0.10880803	2.72271428	0.09628067	-0.31398937	- 0.17054564	1.20400097	-0.99274732	-0.01912070
	1	-0.00958676	0	0	0	0.36984088	-0.03536202	0.61527619	0.14630203	10.02501712	- 6.90246268	0.04175085	-0.01684329
	15	0.01463242	0.01463242	0.01463242	-0.04299412	-0.10938964	0.11532219	-0.23778269	2.28936641	2.43707525	4.68289979	-0.13335552	-0.95880719
	17	0.99954509	0	0	0	5.98924408	0.06263809	-16.29352359	-1.06451953	6.81829629	- 3.53394800	1.24323472	-1.64089851
	18	0	1.07800595	0	0	6.98715683	-0.44220614	11.21847845	-2.04550035	0.51830367	131.95669090	1.76312221	1.12451007
	·4	0.00555023	0	0	0	0.00397249	0.30461678	-0.06785709	0.06087078	--0.18233803	- 0.01336137	0.37031806	0.23618718
B in move No.:	3	0.01798734	0.04437971	0.07460071									
	5	0.02656730	0.05295968	0.13660615	0.08501321	1.26341199	0.63638197						
	11	-0.06734733	0.07043042	0.13359925	0.05242820	1.58137988	0.65471006	2.56494227	0.99260273	- 0.08683532			
	12	-0.02245731	0.16243177	0.16626987	0.09442476	2.51334469	0.58877409	2.83459944	0.78607003	- 0.04561985	0.07951992		
	13	-0.21596752	0.14935465	0.14935466	0.06722218	2.28138607	0.85860046	5.40906583	1.05268608	- 0.21981668	0.06722219	0.92039421	
C in move No.:	2	-0.181950	0	-4.242428									
	4	-1.815979	-1.815979	-3.671656	-2.952842	-380.934473	-330.787773						
	6	291.108111	-8.517857	6.613869	24.078337	57.086056	97.326728	-4192.531659	-582.606293	259.551037			
D	14	0.39370085	0	0	0	0	0	-5.74540777	0	0.35695544	0	-0.63779547	1.00000000
B	15	0.02647735	0.14935465	0.14935466	0.06722218	2.28138607	0.85860046	1.87098707	1.05268608	▨	0.06722219	0.52763347	0.61580986

TABLE (10.11). Continued.

	Sum checks for left members in move No.:						Right members:			
	2	4	6	12	13	14	0	△		
								2	4	6
s=20	198.19	197.39	- 1.71	- 1.71	- 2.81	0.09	3.56491080	- 36.0606	0	0
21	198.19	197.39	194.49	-19.16	-19.16	-19.16	8.79561540	0	0	0
22	198.19	72.84	69.94	69.94	69.94	69.94	14.78511540	-840.8068	0	0
19		258.97537730	270.15290896	270.15290896	270.15290896	270.15290896	16.01202589		0	0
13		2.94345730	-13.64707762	-32.19528890	-32.20583433	-32.17803273	3.53768715		-1106.7801	0
14		2.72271404	2.33445969	3.53846066	2.54571334	2.52659264	1.73268546		- 900.6405	0
1			10.78659495	3.88413226	3.92588312	3.90903983	0.85284213			- 62.797800
15			4.72644114	9.40934094	9.26998541	8.31117822	2.06080713			-701.253870
17			6.81829560	3.28434770	4.52758233	2.88668382	-0.59206893			1769.695876
18				131.95667757	133.71979994	134.84431001	10.49318537			
4					0.37031837	0.60650555	0.34083860			

Working without any sign conditions on the regression coefficients: We will then show how the solution of the problem (10.1) can be obtained by working without any sign condition on the regression coefficients until we reach the last stage.

From the point listed in the last column in the column section 6 in Table (10.2) we now proceed, regardless of the fact that not all the nine regression coefficients on line 7a in the lower left part of Table (10.4) are non-negaive. That is, we make a preference increasing movement with the nine variables 20,21, 22, 19, 13, 14, 17, 1, 15 equal to zero. The corresponding computations are given in the column section 11 of Table (10.9). This leads to a breaking out point where the x_j have the values listed in the last column in the column section 11. According to the standard rule—see line $t = 4$ of Tab. (3.4)—we shall now pick two optimum candidates. As before we pick instead only one since the total work involved will not be any larger than by picking two. This leads through the move number 12—see Table (10.9), (10.10) and (10.11)—to the final attempt No. 13 where the eleven variables

numbers 20, 21, 22, 19, 13, 14, 17, 1, 15, 18, 4 ... (10.12)

are zero and all the d_k are zero.

From the solution obtained by the first method we know that the point we have now reached—and listed in the last column in the column section 12 in Table (10.9)— is actually an optimum point, but by the method now followed we have not yet got a sufficient criterion for the optimality of the point reached because not all the eleven coefficients on the line 13 of Table (10.11) are non-negative. These coefficients are now the B of (7.9).

Looking back we see by the same reasoning as in the first method that number 17 is a black sheep. In the expression for the preference function now obtained —compare (7.9) —we therefore wish to express number 17 in term of eleven other variables, namely the ten variables that are left in the set (10.12) when No. 17 is taken out and another variable added so as to retain the total number eleven.

As such a variable to be added we pick the *next* candidate in the priority order in the last breaking out point, namely the variable number 7. In other words we wish to express number 17 in terms of the 11 variables

numbers 20, 21, 22, 19, 13, 14, 1, 15, 18, 4, 7. ... (10.13)

To do this we must consider the equation between the corresponding twelve boundary vectors, namely the boundary vector for No. 17 and those in (10.13) or — which is the same — the boundary vector for No. 7 and those in (10.12). Compare (7.10) where now $\kappa = 7$. This is done by the previously used 11×11 system and with a right member consisting of the moments defined through the newly added boundary vector No. 7. See (7.11) and (7.12) with $\kappa = 7$. Doing this we find the set of twelve D coefficients listed on line 14 in the lower left part of Table (10.11). Using these coefficients to express the variable No. 17 in terms of those indicated in (10.13) — compare (7.13)—and inserting this expression for No. 17 in the expression for the preference function—compare (7.9)—we get the set of eleven B coefficients listed on line 15 in the lower part of Table (10.11). Since all these are non-negative we know that an optimum point has been reached. We also see that the optimum region is one dimensional and can be expressed in any of the two forms indicated in (10.5)—(10.8).

This example shows that we will as a rule actually reach an optimum point quicker by going ahead without paying any attention to the sign of the regression coefficients, but in return we have to make a transformation at the end if we are absolutely set on getting a sure optimality criterion. When a successful transformation is found so immediately as in this example, it will probably pay to go ahead regardless of the signs of the B-coefficients, but in more complicated cases where a high degree of freedom in the optimum region is to be expected and we are set on getting an optimality criterion, it may be safest to work with non-negative regression coefficients throughout.

11. Illustration of the reversal technique

In the example described in § 10 of the present paper we did not need to use the technique developed in § 9. The numerical material in the example can nevertheless be used to illustrate how the formulae of § 9 work.

Consider for instance the point listed in the last column of column § 6 of Table (10.2) Here the following nine variables

$$20, 21, 22, 19, 13, 14, 1, 15, 17 \qquad \dots \quad (11.1)$$

are zero.

The moment matrix for the boundary vectors of these variables are given in the first nine rows and columns of Table (10.3), and the corresponding matrix solution table is given by the first nine rows and columns of Table (10.4).

The set (11.1) we now consider as the set $a = \alpha, \beta \dots \gamma$ within which the formulae (9.1) and 9.2) are to be applied. We let the set $\theta \dots \delta$ be the set of the three variables

$$1, 15, 17. \qquad \dots \quad (11.2)$$

To determine the nine numbers C'_s $(s = \alpha, \beta \dots \gamma)$, we use (5.3) with all the Δx_s equal to zero except Δx_s for $s = 13$ and $s = 14$, these two Δx_s being put equal to minus the values which the variables Nos. 13 and 14 had in the point from which we started when we wanted to force Nos. 13 and 14 down to zero. This is the point listed in the last column of column section 3 in Table (10.2). In other words we put

$$\begin{aligned} \Delta x_{13} &= -1106.7801 \\ \Delta x_{14} &= -1021.0669. \end{aligned} \qquad \dots \quad (11.3)$$

To compute the numbers C'_s under these conditions is an easy matter because the first nine rows and columns of Table (10.4) have been computed. The result is given on the first line in Table (11.4).

TABLE (11.4). ILLUSTRATION OF THE USE OF THE TECHNIQUE OF SECTION 9.

	s=20	21	22	19	13	14	1	15	17
C'_s	-105.643766	-1.935616	-12.664797	-17.072793	-631.283421	-351.543053	901.027409	147.972876	-95.113408
C^1_s	-2.68356216	0.04550475	0.01151729	-0.05408250	0.35136164	-1.26329507	39.38586322	2.40047400	-2.35255240
C^{15}_s	-0.17116544	0.00790334	-0.01024472	-0.02807007	-0.11051708	-0.02512556	2.40047400	0.60300222	-0.15612691
C^{17}_s	0.16661973	-0.00253122	0.00066853	0.00509160	0.01732396	0.07437287	-2.35255233	-0.15612690	0.14666420
C_s	-1.815999	-1.815971	-3.671672	-2.952867	-380.934230	-330.787728	0	0	0
B_s	0.99561562	0.04144721	0.17209648	0.15980634	2.17313214	0.96500944	-11.86010261	-1.25814957	0.86725465
A_s	0.02656738	0.05295965	0.13660623	0.08501332	1.26341130	0.63638162	0	0	0

Similarly to determine C^1_s we use (5.3) but now with all the $\Delta x_s = 0$ except $\Delta x_1 = 1$. Further C^{15}_s is determined by (5.3) with all the $\Delta x_s = 0$ except $\Delta x_{15} = 1$, and similarly for C^{17}_s. The result is given on the lines Nos. 2, 3, 4 in Table (11.4).

Next the system (9.1)—which is now of order 3 —is solved, giving

$$\Delta x_1 = 377.484406, \quad \Delta x_{15} = -77.206269, \quad \Delta x_{17} = 6685.195685. \qquad \dots \quad (11.5)$$

By (9.2) this gives the values C_s listed on line 5 in Table (11.4). These values are—apart from rounding errors in the last decimal places—identical with the values C listed on the next to last line in Table (10.4).

We can also illustrate the method (9.7)—(9.11). The solutions A_s^1, A_s^{15}, A_s^{17} of (9.8) are identical with the solutions C_s^1, C_s^{15}, C_s^{17} listed on lines Nos. 2, 3, 4 of Table (11.4). Hence the matrix of the system (9.10)—now of order 3—is the same as the one we just used to get (11.5), but B_s in (9.10) is not the same as C'_s in (9.2). B_s is obtained by putting all the Δx_s in (9.8) equal to zero except $\Delta x_r = M_{or}$ for $r = 0 \ldots \delta$. The result is given on the next to last line in Table (11.4). By (9.10) this gives

$$\Delta x_1 = -1.21293526 \qquad \Delta x_{15} = 0.47838524 \qquad \Delta x_{17} = -24.85991535. \qquad (11.6)$$

Finally by (9.9) we get the values of A listed on the last line of Table (11.4). These values are—apart from rounding errors in the last decimal places—identical with the values B on the line marked 5 in Table (10.4).

12. THE REGRESSIONALLY TRANSFORMED PROBLEM

Once the equations have been brought over into *any* basis form, the number of equations does not have a very great influence on the amount of work involved in the multiplex method. It is primarily a large number of degrees of freedom that will entail a heavy computational cost. In many practical problems, not least in macroeconomic planning, the number of degree of freedom, although great from the viewpoint of desk machine computation, will be moderate from the viewpoint of automatic computors, while the number of equations may be extremely large. Furthermore the situation will often be such that the way in which the equations can be brought over into *some* basis form will follow without too much difficulties from the already established theory and practice of the concrete problem at hand. For instance, there may be a certain number of *parameters of action* or *steering parameters*, which will determine the constellation of the system and the way these steering parameters—to a linear approximation—determine a multitude of variables describing the system, may be known. To such cases the multiplex method is well adapted. Even if some basis form of the original equations is not already available, it may be reached through a hierarchic order of *partioning* of the equations or similar devices. Very seldom will all the variables occur in all the equations.

In the opposite case, i.e., when the number of degrees of freedom is very large and the number of equations moderate, we can transform the problem into a linear programming problem where the number *which was originally the number of degrees of freedom now appears as the number of equations and vice versa.*

In essence this is the same as to pass to what is generally called the dual problem. When looked upon from the viewpoint of (4.2) and (7.1) where the B_s are regression coefficients, we may speak of the regressionally transformed problem.

To formulate this problem we first note that in the original problem we can, if we want to, define the preference function in the form

$$f = P_0 + \sum_{j=1}^{n+m} P_j x_j \qquad \ldots \quad (12.1)$$

where the P are any set of $(n+m+1)$ given weights, positive, negative or zero. The passage from this form to (1.5) is obvious, when the equations are given in a basis form.

This being so, consider the following problem. Let B $(j = 1, 2, \ldots, n+m)$ be a set of $n+m$ magnitudes subject to the conditions of : (1) being non-negative, and (2) satisfying the n equations

$$p_k + \sum_{j=1}^{n+m} B_j b_{jk} = 0 \quad (k = u, v \ldots w) \qquad \ldots \quad (12.2)$$

where the p_k and b_{jk} are the coefficients in the original problem. Compare (1.3) and (1.5). To determine that or those sets of values of the magnitudes B_j which satisfy these two sets of conditions and maximize the linear function

$$F = \sum_{j=1}^{n+m} B_j b_{j0} \qquad \ldots \quad (12.3)$$

where the b_{j0} $(j = 1, 2, \ldots, n+m)$ are the constant terms—positive, negative or zero—in the basis form of the equations, i.e. in (1.3).

This obviously is a well defined linear programming problem where the number of variables is the same as in the original problem, namely $(n+m)$, but the number of *equations* is now n and the *number of degrees of freedom* is now m.

We know that the problem now formulated has at least one solution where at least m of the B_j are zero, hence at most n of them different from zero (and then necessarily positive). Inserting these in (4.2) we find that all the n d_k are zero, hence the original preference function f can be written in the form (7.1) with all the B_s non-negative. We therefore have a solution of the original problem.

Thus, a large number of equations need not worry us, provided we can in *some* way without prohibitive work bring the equations over into *any* basis form, nor need a large number of degrees of freedom do so, if in the regressionally transformed problem we can in *some* way bring the equations over into *any* basic form.

13. SOLVING LINEAR EQUATIONS AND INVERTING MATRICES WHEN HIGH SPEED MEMORY IS LIMITED

The following are explicit formulae that can be used in the case where the capacity of the automatic machinery is not large enough to handle the whole problem in one stroke. The essence of this method is that one proceeds step by step as one would with a desk machine and a work sheet, eliminating, however, not *one single* variable at each step, but eliminating *as many* variables at a time as is determined by the order of an inverse which the machine can handle directly in one stroke. While each step in the ordinary Gaussian elimination algorithm involves *one division*, i.e., the formation of one reciprocal number (and several multiplications), each step in the extended algorithm involves the formation of one inverse matrix (and some matrix multiplications). This procedure may, for instance, be used when we want to work the equations of a linear programming problem over into a basis form and there is available automatic machinery with a capacity large enough to handle a good sized inversion but not large enough to make the whole of the desired inversion in one stroke.

Let

$$a_{i0}+\sum_{j=1}^{n+m} a_{ij}x_j = 0 \qquad (i = 1, 2, \ldots, m) \qquad \ldots \quad (13.1)$$

be m linearly independent equations connecting the $n+m$ variables $x_1, x_2, \ldots, x_{n+m}$, the a being given constants.

We pick a set of ν variables ($\nu < m$), let it be Nos. $\alpha \ldots \gamma$, and also pick a set of ν equations, let it be Nos. $\alpha' \ldots \gamma'$, and use these equations to express the ν variables in terms of the other $n+m-\nu$ variables. To do this we consider the matrix

$$\begin{pmatrix} a_{\alpha'\alpha} & \cdots & a_{\alpha'\gamma} \\ \cdots & \cdots & \cdots \\ a_{\gamma'\alpha} & \cdots & a_{\gamma'\gamma} \end{pmatrix} \text{ and its inverse } \begin{pmatrix} a^{-1}_{\alpha\alpha'} & \cdots & a^{-1}_{\alpha\gamma'} \\ \cdots & \cdots & \cdots \\ a^{-1}_{\gamma\alpha'} & \cdots & a^{-1}_{\gamma\gamma'} \end{pmatrix}. \qquad \ldots \quad (13.2)$$

In terms of this inverse, we have

$$x_r = b^{(1)}_{r0} + \sum_{j=k,\,2..)\alpha..\gamma(..n+m} b^{(1)}_{rj} x_j \quad (r = \alpha \ldots \gamma) \qquad \ldots \quad (13.3)$$

where

$$b^{(1)}_{rj} = -\sum_{r'=\alpha'..\gamma'} a^{-1}_{rr'}\, a_{r'j} \quad \begin{pmatrix} r = \alpha \ldots \gamma \\ j = 0, 1 \ldots)\alpha \ldots \gamma(\ldots n+m \end{pmatrix}. \qquad \ldots \quad (13.4)$$

Inserting (12.6) in all the equations (12.4) except the equations Nos. $\alpha' \ldots \gamma'$, we get

$$a^{(1)}_{i0} + \sum_{j=1,2..)\alpha..\gamma(..n+m} a^{(1)}_{ij}\, x_j = 0 \quad (i = 1, 2 \ldots)\alpha' \ldots \gamma'(\ldots m) \qquad \ldots \quad (13.5)$$

where

$$a^{(1)}_{ij} = a_{ij} + \sum_{r=\alpha..\gamma} a_{ir}\, b^{(1)}_{rj} \quad \begin{pmatrix} i = 1, 2 \ldots)\alpha' \ldots \gamma'(\ldots m \\ j = 0, 1 \ldots)\alpha \ldots \gamma(\ldots m+n \end{pmatrix}. \qquad \ldots \quad (13.6)$$

We next pick a new set of variables, now μ in number, ($\mu \leqslant m-\nu$), let it be Nos. $\theta \ldots \delta$, and also pick μ equations in (13.5), let it be Nos. $\theta' \ldots \delta'$. None of the variables Nos. $\theta \ldots \delta$ occur in the set $\alpha \ldots \gamma$, and none of the equations Nos. $\theta' \ldots \delta'$ occur in the set $\alpha' \ldots \gamma'$. Through the equations Nos. $\theta' \ldots \delta$ in (13.5) we express the variables Nos. $\theta \ldots \delta$ in terms of the other $n+m-\nu-\mu$ variables. To do this we consider the matrix

$$\begin{pmatrix} a^{(1)}_{\theta'\theta} & \cdots & a^{(1)}_{\theta\delta} \\ \cdots & \cdots & \cdots \\ a^{(1)}_{\gamma'\theta} & \cdots & a_{\gamma'\delta} \end{pmatrix} \text{ and its inverse } \begin{pmatrix} a^{-1(1)}_{\theta\theta'} & \cdots & a^{-1(1)}_{\theta\delta'} \\ \cdots & \cdots & \cdots \\ a^{-1(1)}_{\delta\theta'} & \cdots & a^{-1(1)}_{\delta\delta'} \end{pmatrix}. \qquad \ldots \quad (13.7)$$

In terms of this inverse we have

$$x_r = b^{(2)}_{ro} + \sum_{j=1,2..)\alpha..\gamma,\,\theta..\delta(..n+m} b^{(2)}_{rj}\, x_j \quad (r = \theta \ldots \delta) \qquad \ldots \quad (13.8)$$

where

$$b^{(2)}_{rj} = -\sum_{r'=\Theta'..\delta'} a^{-1(1)}_{rr'}\, a^{(1)}_{r'j} \quad \begin{pmatrix} r = \theta \ldots \delta \\ j = 0, 1 \ldots)\alpha \ldots \gamma, \theta \ldots \delta\ (\ldots n+m) \end{pmatrix} \qquad \ldots \quad (13.9)$$

Inserting (13.8) in the right member of (12.6) we get

$$x_r = b_{r0}^{(2)} + \sum_{j=1,\,2..)\alpha..\gamma,\theta..\delta(..n+m} b_{rj}^{(2)}\, x_j \quad (r = \alpha \ldots \gamma) \qquad \ldots \quad (13.10)$$

where

$$b_{rj}^{(2)} = b_{rj}^{(1)} + \sum_{s=\theta..\delta} b_{rs}^{(1)}\, b_{sj}^{(2)} \quad \begin{pmatrix} r = \alpha \ldots \gamma \\ j = 0,1\ldots)\alpha\ldots\gamma,\theta\ldots\delta(\ldots n+m \end{pmatrix} \ldots \quad (13.11)$$

Through (13.10) and (13.8) the $\nu+\mu$ variables Nos. $\alpha \ldots \gamma$, $\theta \ldots \delta$ are expressed in terms of the remaining $n+m-\nu-\mu$ variables.

In this way we can continue until we are left with only n variables in the right member. The equations are then brought over into a basis form.

A possibility that might be considered is to express all the variables linearly in *some other* set of n parameters $t_1, t_2 \ldots t_n$. The whole programming theory could easily be expressed in this way. Practical experience will decide if any such transformation will pay.

Paper received : October, 1955.

SOLDIERS WITH UNUSUALLY LOW BLOOD PRESSURE

By N. T. MATHEW
Army Headquarters, India

During the course of an anthropometric survey conducted by the Army Statistical Organisation in 1952, measurements of blood pressure were recorded along with other physical characteristics on a fairly large number of soldiers. When the data was being processed, a number of remarkably low diastolic blood pressures were discovered. These records were naturally viewed with suspicion. To find out whether these were due to errors of observation, it was decided to carry out a reliable check of the measurements. It was not practicable to trace all the subjects for whom low pressures were recorded as they were spread about all over India and had undergone many transfers during the three years that had elapsed. The resurvey was therefore restricted to personnel still available in units located near about Delhi.

In all twentyseven soldiers were made available for the resurvey. Measurements were taken on 27 October, 1955, 28 October, 1955 and 1 November, 1955 by Maj. R. S. Hoon, AMC, a specialist deputed for this purpose and also by Capt. D. N. Bhattacharya, AMC, who made the original measurements.

The technique of measurement was the standard one using a sphygmomanometer. Capt. Bhattacharya used for his original measurements in 1952 an instrument called "Erkameter". This same instrument was used for the 1955 measurements also. Just prior to the resurvey this instrument was tested against two others called "Britometer" and "Baumanometer". The test measurements were made on the left arm of a medical officer. The subject was in sitting position with his arm extended across a table on which the instrument was placed. Capt. D. N. Bhattacharya and Capt. S. K. Dhawan took measurements using the three instruments in turn. While the experiment was about to start one of the medical men suggested that to be quite correct the base of the manometer should be at the level of the subject's heart, and not on the table which was about 8″ below this level. It was decided to test this theory also. So measurements were taken at two levels by each observer using each instrument. The measurements in the order in which they were obtained are given in Table 1. The physical particulars of the subject are: age 38 years, height 69″, weight 168 lbs.

TABLE 1. BLOOD PRESSURE MEASUREMENTS TAKEN AT 26 GENERAL HOSPITAL ON 27 OCTOBER 1955

time 1	observer 2	instrument 3	level 4	systolic pressure 5	diastolic pressure 6
0902	1	*A*	low	122	84
0907	1	*B*	low	116	80
0910	1	*C*	low	114	80
0918	2	*A*	low	110	80
0921	2	*A*	high	118	86
0922	1	*A*	high	118	80
0926	2	*B*	low	110	82
0930	2	*B*	high	114	84
0932	1	*B*	high	112	80
0935	2	*C*	low	110	80
0937	2	*C*	high	110	84
0939	1	*C*	high	112	80

The figures in Table 1 show incidentally that even within a period of half an hour the readings can vary from 80 to 86 for diastolic pressure and 110 to 122 for systolic pressure. Comparisons between observers, between instruments and between levels based upon the above data are presented in Table 2.

TABLE 2. SHOWING SIGNIFICANCE OF DIFFERENCES IN MEAN BLOOD PRESSURES

comparisons (1)	*n* (2)	systolic pressure (*mm*) (3)	diastolic pressure (*mm*) (4)
observers			
1. Bhattacharya	6	115.7	80.7
2. Dhawan	6	112.0	82.7
F test of difference	—	not significant	significant at 5%
instruments			
A Erkameter	4	117.0	82.5
B Britometer	4	113.0	81.5
C Baumanometer	4	111.5	81.0
F test of differences	—	not significant	not significant
levels			
high	6	114.0	82.3
low	6	113.7	81.0
F test of difference	–	not significant	not significant

The mean squares for different factors when expressed as a ratio of the corresponding error mean square are less than the magnitude required for significance. In diastolic pressure the difference between observers (though not significant at the 1% level) is significant at the 5% level. However, the actual difference between the two means is only 2 millimeters which is of little consequence from a practical point of view.

Having examined the preliminary questions, we now pass on to the measurements on subjects who had unusually low blood pressure. All these measurements were made with the 'Erkameter' which we have designated as Instrument *A* in the foregoing experiment. The subjects had at least an hour's rest before measurement. The sitting posture was adopted in every measurement. Measurements were made on both the right arm and the left arm of each subject by Hoon as well as by Bhattacharya. The measurements are given in Table 3 in the order in which they were taken. The room temperature at 1109 hrs. on 27 October was 76°F., on 28 October it was 68°F at 0920 hrs. and 74°F at 1115 and on 1 November 1955 it was 69°F at 0845.

TABLE 3. SHOWING MEASUREMENTS OF BLOOD PRESSURE MADE BY R. S. HOON IN 1955 AND D. N. BHATTACHARYA IN 1955 AND 1952

	in 1955			by R. S. Hoon (1955)						by D. N. Bhattacharya (1955)						by D. N. Bhattacharya (1952)			
				right arm			left arm			right arm			left arm			right arm		left arm	
sl. no.	age in years	height in inches	weight in lbs	time	S.P.	D.P.	time	S.P.	D.P.	time	S.P.	D.P.	time	S.P.	D.P.	S.P.	D.P.	S.P.	D.P.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
27 October 1955																			
1	26	66	112	1013	85	60	1016	80	55	1019	86	50	1022	88	50	102	50	90	40
2	25	72	130	1026	90	52	1030	88	50	1035	84	48	1039	84	44	86	50	80	40
3	33	68	136	1045	100	58	1049	94	52	1048	92	58	1052	90	62	94	50	90	40
4	22	69	156	1109	90	60	1112	80	55	1121	90	48	1114	86	48	100	54	90	54
5	22	69	156	1109	90	60	1112	80	55	1121	90	48	1114	86	48	94	44	90	54
6	31	67	128	1138	90	60	1139	88	60	1135	88	44	1140	90	50	90	54	96	56
7	23	70	146	1153	88	55	1155	94	58	1150	88	54	1153	88	60	94	54	90	40
8	31	66	126	1203	90	65	1206	90	55	1200	98	60	1204	96	60	88	50	90	40
28 October 1955																			
9	23	67	126	0929	98	75	0934	94	62	0926	100	68	0933	94	70	90	48	84	40
10	40	69	138	0942	100	72	0949	88	64	0940	98	60	0947	86	58	90	60	90	60
11	38	64	110	0951	88	58	0957	84	54	0955	88	62	1000	84	60	88	48	99	48
12	25	66	110	1005	84	55	1010	82	50	1008	84	48	1015	74	46	88	48	88	48
13	24	67	132	1025	84	52	1027	84	52	1018	90	46	1023	90	54	88	40	88	40
14	23	67	124	1036	90	65	1038	88	62	1032	94	64	1035	90	64	100	50	90	40
15	24	68	126	1040	88	65	1042	88	62	1045	90	58	1047	86	48	92	50	84	40
16	23	71	148	1051	100	58	1053	102	64	1058	96	68	1101	90	68	96	46	90	50
17	30	67	126	1115	88	54	1113	90	58	1110	86	56	1106	84	60	100	40	108	54
18	25	65	136	1122	86	56	1119	88	54	1129	90	50	1126	86	50	100	64	90	60
19	29	66	126	1140	98	66	1113	98	68	1138	98	68	1132	96	60	84	40	80	40
20	32	64	116	1148	102	70	1141	105	65	1150	100	64	1145	98	68	104	60	90	60
1 November 1955																			
21	33	70	124	0815	84	56	0823	80	54	0813	78	58	0820	78	54	88	60	86	56
22	24	68	120	0832	106	56	0837	100	64	0830	100	82	0835	104	74	100	60	90	60
23	24	71	124	0846	94	54	0851	84	54	0845	98	66	0849	86	64	100	50	90	40
24	29	68	126	0859	84	44	0907	78	54	0858	78	50	0906	78	60	76	40	80	40
25	33	69	128	0915	90	62	0925	84	58	0911	86	72	0918	78	62	100	50	90	40
26	32	66	116	0934	84	62	0940	90	66	0932	84	62	0937	88	68	104	70	90	60
27	27	67	128	0945	94	66	0950	92	56	0943	92	80	0948	88	62	104	54	90	60

Comparison of the average values of measurements obtained in 1955 are given in Table 4. The average age is 27.9 years, height 67.2″ and weight 127.0 lbs.

TABLE 4. TESTS OF SIGNIFICANCE OF DIFFERENCES IN BLOOD PRESSURE READINGS

comparisons (1)	means (2)	(3)	t (4)	significance (5)
(a) between observers, 1955				
1. right arm, S.P.	91.5	90.7	0.98	—
2. right arm, D.P.	59.9	58.9	0.52	—
3. left arm, S.P.	89.2	87.9	1.56	—
4. left arm, D. P.	58.0	58.3	−0.23	—
(b) between right and left arm				
1. Hoon; S.P.; 1955	91.5	89.2	2.63	sig.
2. Bhattacharya; S.P.; 1955	90.7	87.9	3.45	very sig.
3. Hoon; D.P.; 1955	59.9	58.0	1.75	—
4, Bhattacharya; D.P.; 1955	58.9	58.3	0.49	—
5. Bhattacharya; S.P.; 1952	93.7	88.9	3.65	very sig.
6. Bhattacharya; D.P.; 1952	50.9	48.1	2.00	—
(c) between 1952 readings of Bhattacharya and				
1. Hoon; S.P.; 1955; right arm	91.5	93.7	−1.30	
2. Bhattacharya; S.P.; 1955; right arm	90.7	93.7	−1.73	—
3. Hoon; D.P.; 1955; right arm	59.9	50.9	5.22	very sig.
4. Bhattacharya; D.P.; 1955; right arm	58.9	50.9	3.48	very sig.
5. Hoon; S.P.; 1955; left arm	89.2	88.9	0.14	—
6. Bhattacharya: S.P.; 1955; left arm	87.9	88.9	−0.65	—
7. Hoon; D.P.; 1955; left arm	58.0	48.1	5.92	very sig.
8. Bhattacharya; D.P.; 1955; left arm	58.3	48.1	5.53	very sig.

Values of 't' for testing the significance of the differences between the different comparable series from Table 3 are also given in col. 4 of Table 4.

The following conclusions emerge from Table 4.

(a) There is no significant difference between the measurements made by Hoon and those made by Bhattacharya on the same subjects. We may, therefore, accept the results of the 1952 survey carried out by Bhattacharya as free from any significant error of measurements.

(b) The averages for the right arm appear to be higher than the corresponding averages for the left arm, though the difference is significant only for systolic pressure. In a personal discussion with Lt. Col. G. C. Tandon and Lt. Col. P. T. Joseph about this result, they pointed out that the difference might be due to readings on the right hand being taken first. It is well known that the initial rise of blood pressure due to the nervousness caused by the measuring operation subsides with time, and hence the above is quite a plausible explanation for the results observed. From Table 3 it can be seen that the observation on the right arm preceded that on the left arm on most subjects. It is not possible, therefore, to separate the effect of the initial nervousness from the real difference, if any, between right and left arms. However, it happened that on serial numbers 4, 17, 18, 19 and 20, readings were taken on the left arm before readings on the right arm. These readings are reproduced in Table 5.

TABLE 5. BLOOD PRESSURE READINGS ON THE LEFT ARMS BEFORE THE READINGS ON THE RIGHT ARMS

	sl. no. (1)	right arm (2)	left arm (3)	difference (4)
Bhattacharya				
	4	90	86	4
	17	86	84	2
	18	90	86	4
	19	98	96	2
	20	100	98	2
Hoon				
	17	88	90	-2
	18	86	88	-2
	19	98	98	0
	20	102	105	-3

From Table 5 it will be seen that the measurements obtained by Bhattacharya show higher readings on the right arm, even when taken after the readings on the left arm. The measurements taken by Hoon, however, show the opposite tendency. Taken as a whole, these nine differences are not statistically significant. It is possible, therefore, that the differences we have observed between right arm and left arm are entirely due to the order of measurement. The point is worth further investigation in view of the consistency of the differences in the five pairs of measurements by Bhattacharya.

(c) The comparisons between the readings obtained in 1955 and those obtained three years earlier by Bhattacharya indicate the possibility of a real increase of the diastolic pressure in the subjects measured. The 1955 readings of diastolic pressure by Hoon as well as by Bhattacharya are significantly higher than the corresponding 1952 readings.

It appears reasonable from all these to conclude that the low blood pressures observed in 1952 by Bhattacharya really indicate the existence of normal active men in the army with blood pressure much lower than what is usually regarded as normal. In Table 6 we give the distribution of the 1952 readings of systolic pressure and in Table 7 the corresponding figures for diastolic pressure as recorded by Bhattacharya. A detailed analysis of this data in relation to age, height, weight, community, province of origin etc., is in progress, and a separate publication on the results will follow.

From Table 6 it will be seen that systolic pressure varies among different persons from 75 mm to over 180 mm. However, if we leave out the small frequencies the range from 75 to 134 mm includes well over 99% of all the cases. Similarly from Table 7 it can be seen that the range from 40 to 84 mm will cover 99% of all the diastolic pressures.

TABLE 6. DISTRIBUTION OF SYSTOLIC BLOOD PRESSURE OF SOLDIERS OF THE INDIAN ARMY (BHATTACHARYA 1952)

	right arm			left arm		
pressure in *mm*	frequency	percentage of total	cumulative percentage	frequency	percentage of total	cumulative percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
75– 84	67	2.5	2.5	83	3.4	3.4
85– 94	423	16.4	18.9	448	19.1	22.5
95–104	861	33.0	51.9	779	33.4	55.9
105–114	778	30.0	81.9	663	28.4	84.3
115–124	353	13.8	95.7	284	12.0	96.3
125–134	93	3.5	99.2	74	3.2	99.5
135–144	18	.7	99.9	7	.3	99.8
145–154	3	.1	100.0	4	.2	100.0
155–164	1	.0	100.0	—	—	100.0
165–174	—	—	100.0	1	.0	100.0
175–185	1	.0	100.0	1	.0	100.0
total	2598	100.0	100.0	2344	100.0	100.0

TABLE 7. DISTRIBUTION OF DIASTOLIC BLOOD PRESSURE OF SOLDIERS OF THE INDIAN ARMY (BHATTACHARYA 1952)

	right arm			left arm		
pressure in *mm*	frequency	percentage of total	cumulative percentage	frequency	percentage of total	cumulative percentage
(1)	(2)	(3)	(4)	(5)	(6)	(7)
30– 34	8	.3	.3	7	.3	.3
35– 44	404	15.4	15.7	422	18.1	18.4
45– 54	937	35.8	51.5	877	37.6	56.0
55– 64	764	29.1	80.6	663	28.3	84.3
65– 74	392	16.0	96.6	310	13.1	97.4
75– 84	77	2.9	99.5	57	2.3	99.7
85– 94	13	.5	100.0	7	.3	100.0
95–104	1	.0	100.0	1	.0	100.0
total	2596	100.0	100.0	2344	100.0	100.0

The percentage distributions of the systolic and diastolic pressures given respectively in column 3 of Table 6 and column 3 of Table 7 are represented graphically in Charts 1 and 2.

Chart 1. Distribution of systolic pressure of soldiers of the Indian Army 1952

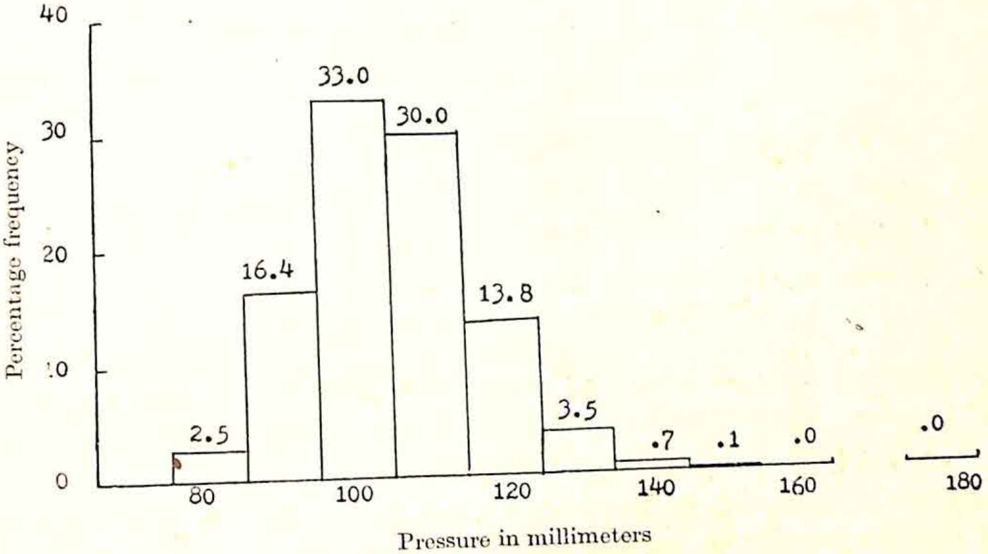
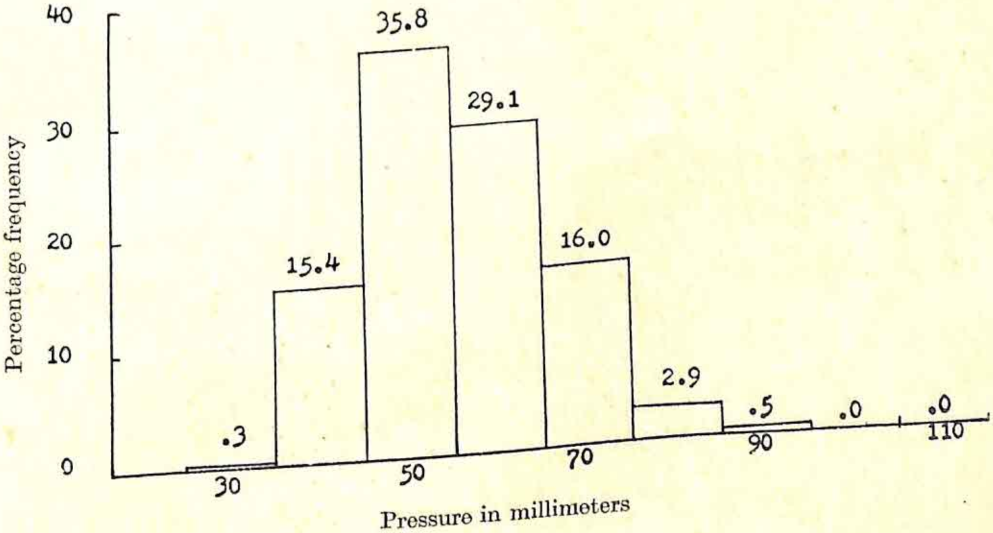


Chart 2. Distribution of diastolic pressure of soldiers of the Indian Army 1952



Means and standard deviations of these measurements are given in Table 8.

TABLE 8. MEANS AND STANDARD DEVIATIONS OF BLOOD PRESSURE OF SOLDIERS OF THE INDIAN ARMY MEASURED IN 1952

characteristic	N	mean (mm)	standard deviation (mm)
(1)	(2)	(3)	(4)
right arm S.P.	2598	106.0	11.6
left arm S.P.	2340	104.5	11.6
right arm D.P.	2596	57.0	11.6
left arm D.P.	2340	55.9	10.3

The above averages will be regarded as low by the standards (if there be any) that are now being used in practice. The lowness is due to the inclusion of a large proportion of unusually low values both in systolic and in diastolic pressures.

I am thankful to Capt. D. N. Bhattacharya and Lt. Col. R. S. Hoon for their work in collecting the data used above, and to a number of other officers of the AMC for advice. Particular mention must be made of Maj. M. C. Sanyal who went through the manuscript.

Paper received : April, 1957.

UNBIASSED MINIMUM VARIANCE ESTIMATION IN A CLASS OF DISCRETE DISTRIBUTIONS

By JOGABRATA ROY and SUJIT KUMAR MITRA
Indian Statistical Institute, Calcutta

The problem of estimation of parameters involved in discrete distributions like the Binomial, Poisson and Negative Binomial and their truncated forms have been considered by various authors. Most of them, e.g., Fisher (1936, 1941), Haldane (1941), Finney (1949, 1955) David and Johnson (1952) discuss computational aspects of estimation by maximum likelihood while a few others e.g., Plackett (1953), Moore (1952, 1954), Rider (1953, 1955) give other simpler but inefficient methods of estimation which at times lead to unbiassed estimates.

In this paper, for a wide class of discrete distributions involving one unknown parameter the uniformly minimum variance unbiassed (UMVU) estimate of the parameter is derived and the UMVU estimate of its variance is also obtained. It is shown that results for distributions truncated at zero can be obtained from those for the complete disbribution by a difference operation. The negative binomial and the Poisson distributions are treated as special cases. Tables of the UMVU estimate are given for the Poisson distribution truncated at zero, for sample size upto ten.

1. UMVU ESTIMATES IN A CLASS OF DISCRETE DISTRIBUTIONS

Consider the following discrete probability distribution defined by Noak (1950).

$$\text{prob}\ \{X = x\} = a(x)\theta^x/f(\theta),\ x = 0, 1, 2, \ldots, \qquad \ldots \quad (1.1)$$

where $\theta > 0$ is an unknown parameter, $a(x) > 0$ does not involve θ and

$$f(\theta) = \sum_{x=0}^{\infty} a(x)\theta^x. \qquad \ldots \quad (1.2)$$

Without loss of generality, we shall assume $a(0) = 1$. The Poisson, Negative Binomial and Logarithmic series distributions occur as special cases of the above.

To derive uniformly minimum variance unbiassed (UMVU) estimate for θ on the basis of a random sample $X_i (i = 1, 2, \ldots n)$ of size n from the distribution (1.1) we require the lemmas stated below without proof.

Lemma 1.1: *If* $t_r(x)$ *is defined as*

$$\left.\begin{aligned} t_r(x) &= 0 \quad &\text{for } x < r \\ &= \frac{a(x-r)}{a(x)} \quad &\text{for } x \geqslant r \end{aligned}\right\} \qquad \ldots \quad (1.3)$$

then $E\{t_r(X)\} = \theta^r$.

Lemma 1.2: $T = \sum_{i=1}^{n} X_i$ *is a complete sufficient statistic for* θ *in the sense of Lehmann and Scheffé* (1950).

Lemma 1.3: *The probability distribution of T is given by*

$$\text{prob}\ \{T = t\} = C(t, n)\, \theta^t / f(\theta)^n,\ t = 0, 1, 2, \ldots, \qquad \ldots\ (1.4)$$

where $C(t, n)$ is the coefficient of θ^t in the expansion of $\{f(\theta)\}^n$ and it may be expressed alternatively as

$$C(t, n) = \Sigma \prod_{i=1}^{n} a(x_i),$$

Σ denoting summation over non-negative integral values of $x_1, x_2, \ldots, x_n$ subject to $x_1+x_2+\ldots+x_n = t$.

Defining for any positive integer r,

$$\left.\begin{aligned} u_r(t) &= 0 && \text{for } t < r \\ &= \frac{C(t-r, n)}{C(t, n)} && \text{for } t \geqslant r \end{aligned}\right\} \qquad \ldots\ (1.5)$$

and noting that the probability distribution of T is of the same form as (1.1) it follows from Lemma 1.1 that

$$E\{u_r(T)\} = \theta^r. \qquad \ldots\ (1.6)$$

Incidentally we may note that

$$u_r(t) = \prod_{i=0}^{r-1} u_1(t-i). \qquad \ldots\ (1.7)$$

From (1.6) and the completeness of the sufficient statistic T it follows from Rao-Blackwell Theorem (Blackwell, 1947 and Rao, 1945) that $u_r\ (T)$ is the UMVU estimate of θ^r. The variance of $u_1(T)$, the UMVU estimate of θ is given by

$$V\{u_1(T)\} = E\{u_1(T)\}^2 - \theta^2. \qquad \ldots\ (1.8)$$

Hence we have the following:

Theorem 1.1: *For the distribution* (1.1) *the UMVU estimate for θ is $u_1(T)$ and the UMVU estimate of the variance of $u_1(T)$ is*

$$v(T) = u_1(T) - u_2(T) \qquad \ldots\ (1.9)$$

which because of (1.7) *may be put in the form:*

$$v(T) = u_1(T)\{u_1(T) - u_1(T-1)\} \qquad \ldots\ (1.10)$$

where $u_r\ (T)$ for $r = 1, 2$, are defined by (1.5).

2. UMVU ESTIMATE IN THE CASE OF TRUNCATION ON THE LEFT

Consider the distribution (1.1) truncated on the left at $x = s-1$. The probability distribution can then be written as

$$\text{prob}\ \{X = x\} = a(x)\theta^x / f_s(\theta),\ x = s, s+1, s+2, \ldots, \qquad \ldots\ (2.1)$$

where

$$f_s(\theta) = \sum_{x=s}^{\infty} a(x)\, \theta^x. \qquad \ldots\ (2.2)$$

By an argument exactly similar to the one used in section 1, it follows that on the basis of a random sample $X_i (i = 1, 2, \ldots, n)$ of size n from the distribution (2.1) the UMVU estimate for θ^r is $u_{r,s}(T)$ where

$$\left.\begin{aligned} u_{r,s}(t) &= 0 && \text{for } t < ns+r \\ &= \frac{C_s(t-r,n)}{C_s(t,n)} && \text{for } t \geqslant ns+r \end{aligned}\right\} \quad \ldots \quad (2.3)$$

where $T = \sum_{i=1}^{n} X_i$ and

$$C_s(t,n) = \Sigma_s \prod_{i=1}^{n} a(x_i), \quad \ldots \quad (2.4)$$

Σ_s denoting summation over integral values of $x_1, x_2, \ldots, x_n$ subject to $x_1+x_2 \ldots +x_n = t$ and $x_i \geqslant s$ $(i = 1, 2, \ldots n)$.

The following relation connecting $C_s(t, n)$ can be easily established,

$$C_s(t,n) = \sum_{j=0}^{n-1} (-1)^j \binom{n}{j} \{a(s-1)\}^j C_{s-1}\{t-j(s-1), n-j\}, \quad \ldots \quad (2.5)$$

where we write $C_0(t,n) = C(t,n)$.

For the special case of interest of truncation at zero, that is $s = 1$, we have

$$C_1(t,n) = \sum_{j=0}^{n-1} (-1)^j \binom{n}{j} C(t, n-j), \quad \ldots \quad (2.6)$$

which may be expressed in the alternative form :

$$C_1(t,n) = \Delta^n \, C(t,x)]_{x=0} \quad \ldots \quad (2.7)$$

where Δ is the ordinary difference operator on $C(t, x)$ regarded as a function of the integer x.

In general, introducing the bivariate operator Δ_k defined by

$$\Delta_k \, f(x,y) = f(x+k, y+1) - a(k) f(x,y),$$

we can write (2.5) in the alternative form :

$$C_s(t,n) = \Delta_{s-1}^{n} \; C_{s-1}\{t-n(s-1), 0\}. \quad \ldots \quad (2.8)$$

3. Illustrations

3.1. *Complete Negative Binomial Distribution.* We shall consider the following form of the complete Negative Binomial distribution:

$$\text{prob } \{X = x\} = \frac{(k+x-1)!}{(k-1)! \; x!} \, \theta^x / (1-\theta)^{-k}, \; x = 0, 1, 2, \ldots, \quad (3.1)$$

and take k as a known positive integer as in problems of inverse binomial sampling. It is well-known that the total T of a random sample of size n from this distribution has again the Negative Binomial distribution.

$$\text{prob } \{T=t\} = \frac{(kn+t-1)!}{(kn-1)!\,t!}\,\theta^t/(1-\theta)^{-kn},\ t=0,1,2,\ldots, \qquad \ldots \quad (3.2)$$

so that here

$$C(t,n) = \frac{(kn+t-1)!}{(kn-1)!\,t!}\,. \qquad \ldots \quad (3.3)$$

Hence for the complete distribution we have the following.

Result 3.1: The UMVU estimate of θ on the basis of a random sample of size n from the Negative Binomial distribution (3.1) is

$$T/(kn+T-1), \qquad \ldots \quad (3.4)$$

and the UMVU estimate of the variance of this estimate is

$$\frac{(kn-1)\,T}{(kn+T-1)^2(kn+T-2)}\,, \qquad \ldots \quad (3.5)$$

where T is the total of the sample.

3.2. *Poisson distribution truncated on the left at zero.* For the complete Poisson distribution

$$\text{prob } \{X=x\} = \frac{1}{x!}\,\theta^x/e^\theta, \quad x=0,1,2,\ldots, \qquad \ldots \quad (3.6)$$

it is well known that the distribution of the total T of a random sample of size n is given by

$$\text{prob } \{T=t\} = \frac{n^t}{t!}\,\theta^t/e^{n\theta}, \qquad \ldots \quad (3.7)$$

which is again of the Poisson form. Hence,

$$C(t,n) = \frac{n^t}{t!}\,. \qquad \ldots \quad (3.8)$$

Therefore for a Poisson distribution truncated on the left at zero, with probability law given by

$$\text{prob } \{X=x\} = \frac{1}{x!}\,\theta^x/(e^\theta-1),\ x=1,2,\ldots, \qquad \ldots \quad (3.9)$$

we have from (3.3) and (2.7) $\quad C_1(t,n) = \left.\frac{\Delta^n x^t}{t!}\right]_{x=0}$

which may be written in the usual form:

$$C_1(t,n) = \frac{\Delta^n 0^t}{t!}. \qquad \ldots \quad (3.10)$$

Using (2.3) and the results of Theorem 1.1 we thus have the following:

Result 3.2: On the basis of a random sample of size n from a Poisson distribution truncated on the left at zero as given by (3.9), the UMVU estimate for θ is $u_1(T)$ and the UMVU estimate for the variance of $u_1(T)$ is $V(T) = u_1(T)\{u_1(T)-u_1(T-1)\}$ where

$$\left.\begin{aligned} u_1(t) &= 0 \qquad \text{for } t < n+1 \\ &= \frac{t\Delta^n 0^{t-1}}{\Delta^n 0^t} \quad \text{for } t \geqslant n+1 \end{aligned}\right\}$$

T being the sample total.

Table 2 gives values of $u_1(t)$ correct to three places of decimals for $n = 2, 3, \ldots, 10$. The values of the differences of zero tabulated by Fisher and Yates (1933) may be used for $n > 10$ if T happens to be less than 25.

On the basis of a sample $x_1, x_2, \ldots x_n$ from a Poisson distribution truncated at zero Plackett (1953) suggested an unbiassed estimate

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

where

$$\left.\begin{aligned} y_i &= 0 \quad \text{if } x_i = 1 \\ &= x_i \quad \text{if } x_i > 1. \end{aligned}\right\}$$

The variance of this estimate is

$$V(\hat{\theta}) = \frac{1}{n}\left\{\theta + \frac{\theta^2}{e^{\theta}-1}\right\},$$

The following table (Table 1) shows the efficiency of Plackett's estimate relative to the UMVU estimate for $n = 10$ and several values of θ.

TABLE 1. EFFICIENCY OF PLACKETT'S ESTIMATES RELATIVE TO UMVU ESTIMATE BASED ON A SAMPLE OF SIZE 10

θ	variance of Plackett's estimate	variance of UMVU estimate	efficiency (3)/(2)
(1)	(2)	(3)	(4)
1.0	0.158198	0.151703	0.9589
2.0	0.262607	0.252182	0.9603
3.0	0.347156	0.338585	0.9753
4.0	0.439852	0.424500	0.9875

This shows that Plackett's estimate is highly efficient. For larger sample sizes there is a slight fall in efficiency.

TABLE 2. UMVU ESTIMATE OF THE PARAMETER IN POISSON DISTRIBUTION TRUNCATED AT ZERO (BASED ON A SAMPLE OF SIZE n WHEN THE TOTAL OF THE OBSERVATIONS IS t)

t \ n	2	3	4	5	6	7	8	9	10
2	0	—	—	—	—	—	—	—	—
3	1.000	0	—	—	—	—	—	—	—
4	1.714	0.667	0	—	—	—	—	—	—
5	2.333	1.200	0.500	0	—	—	—	—	—
6	2.903	1.667	0.923	0.400	0	—	—	—	—
7	3.444	2.093	1.300	0.750	0.333	0	—	—	—
8	3.969	2.493	1.646	1.067	0.632	0.286	0	—	—
9	4.482	2.874	1.970	1.360	0.950	0.545	0.250	0	—
10	4.990	3.242	2.278	1.635	1.159	0 786	0.480	0.222	0
11	5.495	3.601	2.574	1.896	1.399	1.011	0.694	0.419	0.200
12	5.997	3.953	2.860	2.146	1.627	1.224	0.9 6	0.622	0.387
13	6.498	4.299	3.139	2.388	1.846	1.427	1.088	0.805	0.564
14	6.999	4.642	3.412	2.623	2.057	1.622	1.272	0.980	0.731
15	7.500	4.983	3.680	2.852	2.262	1.810	1.448	1.147	0.892
16	*	5.321	3.945	3.076	2.461	1.993	1.618	1.309	1.046
17		5.658	4.206	3.297	2.656	2.170	1.783	1.464	1.194
18		5.994	4.466	3.514	2.847	2.343	1.944	1.615	1.338
19		6.329	4.723	3.728	3.034	2.513	2.101	1.762	1.477
20		6.664	4.979	3.940	3.219	2.679	2.254	1.096	1.613
21		6.998	5.233	4.150	3.401	2.843	2.404	2.046	1.746
22		7.332	5.487	4.358	3.582	3.004	2.552	2.184	1.875
23		7.666	5.740	4.565	3.760	3.163	2.697	2.319	2.002
24		7.999	5.992	4.771	3.937	3.320	2.840	2.451	2.127
25		8.333	6.244	4.976	4.112	3.476	2.982	2.582	2.249
26		*	6.495	5.180	4.286	3.630	3.121	2.711	2.370
27			6.746	5.384	4.460	3.783	3.259	2.838	2.489
28			6.997	5:586	4.632	3.934	3.396	2.964	2.606
29			7.248	5.789	4.803	4.085	3.532	3.088	2.722
30			7.498	5.991	4.974	4.235	3.666	3.121	2.836
31			7.749	6.192	5.145	4.384	3.800	3.334	2.950
32			7.999	6.394	5.314	4.532	3.933	3.334	2.950
33			8.249	6.595	5.484	4.679	4.065	3.575	3.173
34			8.499	6.796	5.652	4.826	4.196	3.694	3.284
35			8.750	6.996	5.821	4.973	4.327	3.813	3.393
36			*	7.197	5.990	5.119	4.457	3.931	3.502
37				7.398.	6.158	5.265	4.586	4.049	3.610
38				7.598	6.326	5.410	4.715	4.166	3.717
39				7.798	6.494	5.555	4.844	4.282	3.824
40				7.999	6.661	5.700	4.972	4.398	3.930

---

* For values of t beyond this use t/n.

TABLE 2 (*Contd.*)

t \ n	5	6	7	8	9	10
41	8.199	6.829	5.845	5.100	4.513	4.036
42	8.399	6.996	5.989	5.228	4.628	4.141
43	8.599	7.163	6.133	5.355	4.743	4.246
44	8.799	7.331	6.277	5.482	4.957	4.351
45	9.000	7.498	6.421	5.609	4.971	4.455
46	*	7.665	6.565	5.736	5.085	4.557
47		7.832	6.709	5.862	5.199	4.662
48		7.999	6.852	5.989	5.312	4.765
49		8.165	6.996	6.115	5.425	4.868
50		8.332	7.139	6.241	5.538	4.971
51		8.499	7.282	6.367	5.651	5.073
52		8.666	7.426	6.493	5.763	5.175
53		8.833	7.569	6.619	5.867	5.278
54		8.999	7.712	6.744	5.988	5.379
55		9.166	7.855	6.870	6.101	5.481
56		*	7.998	6.996	6.213	5.583
57			8.141	7.121	6.325	5.684
58			8.284	7.246	6.437	5.786
59			8.427	7.372	6.549	5.887
60			8.570	7.497	6.660	5.988
61			8.713	7.623	6.772	6.089
62			8.856	7.748	6.884	6.190
63			8.999	7.873	6.995	6.291
64			9.142	7.998	7.107	6.392
65			9.285	8.123	7.218	6.492
66			*	8.249	7.330	6.593
67				8.374	7.441	6.694
68				8.499	7.553	6.794
69				8.624	7.664	6.895
70				8.749	7.776	6.995
71				8.784	7.887	7.096
72				8.999	7.998	7.196
73				9.124	8.109	7.296
74				9.249	8.221	7.397
75				9.375	8.332	7.497
76				*	8.443	7.579
77					8.554	7.697
78					8.666	7.798
79					8.777	7.898
80					8.888	7.998

TABLE 2 (*Contd.*)

t \ n	9	10
81	8.999	8.098
82	9.110	8.198
83	9.222	8.299
84	9.333	8.399
85	9.444	8.499
86	*	8.599
87		8.699
88		8.799
89		8.899
90		8.999
91		9.099
92		9.199
93		9.299
94		9.399
95		9.500
96		*

REFERENCES

BLACKWELL, D. (1947): Conditional expectation and unbiassed sequential estimation. *Ann. Math. Stat.*, **18**, 105-110.

DAVID, F. N. AND JOHNSON, N. L. (1952): The truncated Poisson. *Biometrics*, **8**, 175-185.

FINNEY, D. J. (1949). The truncated Binomial distribution. *Ann. Eug.*, **14**, 319 328.

FINNEY, D. J. AND VARLEY, G. C. (1955): An example of the truncated Poisson distribution. *Biometrics*, **11**, 387—394.

FISHER, R. A. (1936): The effects of method of ascertainment upon estimation of frequencies. *Ann. Eug.* **6**, 13-25.

——— (1941): The Negative Binomial Distribution. *Ann. Eug.* **11**, 182-187.

FISHER R A. AND YATES F. (1953): *Statistical Tables for Biological, Agricultural and Medical Research* Fourth Edition, Oliver & Boyd, Edinburgh, 1953.

HALDANE. J. B. S. (1941): The fitting of Binomial distribution *Ann. Eug.* **11**, 179-181.

LEHMANN, E. L. AND SCHEFFÉ, H. (1950): Completeness, Similar Regions and Unbiassed estimation. Part I. *Sankhyā*, **10**, 305-340.

MOORE, P. G. (1952): The estimation of the Poisson parameter from a truncated distribution. *Biometrika*, **39**, 247-251.

——— (1954): A note on truncated Poisson distribution. *Biometrics*, **10**, 402-406.

NOAK, A. (1950): A class of random variables with discrete distributions. *Ann. Math. Stat.*, **21**, 127-132.

PLACKETT, R. L. (1953): The truncated Poisson distribution. *Biometrics*, **9**, 485-488.

RAO, C. R. (1945): Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, **37**, 81-91.

RIDER, P. R. (1953): Truncated Poisson distributions. *J. Amer. Stat. Ass.*, **48**, 826-830.

——— (1955): Truncated Binomial and Negative Binomial distributions. *J. Amer. Stat. Ass.*, **50**,

Paper received : January, 1957.

ORDERED AND UNORDERED ESTIMATORS IN SAMPLING WITHOUT REPLACEMENT

M. N. MURTHY

Indian Statistical Institute, Calcutta

1. Introduction

In the last decade the probem of sampling with varying probabilities without replacement has received considerable attention and a number of estimators have been proposed. Das (1951) and Des Raj (1956) have given some ordered estimators, that is, estimators which take into account the order in which the units are drawn. In this paper it is shown that corresponding to any biased or unbiased ordered estimator there exists an unordered estimator which is more efficient than the former. The technique of improving the ordered estimators by unordered ones is also explained. This method is applied to the set of estimators given by Das and to one set of Des Raj's estimators which provide unbiased estimates of the population total and also to the unbiased variance estimators considered by them. ($\bar{x}_{si}$, v_{si}, $\bar{x}'_{si}$, v'_{si}, of Sections 3 and 4). Compared to other known estimators the estimator of Des Raj has two remarkable properties : (i) $\bar{x}_{si}$ is more efficient than the corresponding estimator for sampling with replacement as shown by Roy Choudhury (1956) and (ii) v_{si} is non-negative. It is shown that the technique of unordering preserves a fortiori the first property and retains the second property in two particular cases. In sampling the first unit with varying probability and the rest with equal probability without replacement, unordering of Das' estimator yields the familiar unbiased ratio estimator.

2. Unordered estimator

In sampling n units without replacement from a finite population of N units, there will be $\binom{N}{n}$ unordered samples (s). Corresponding to any unordered sample (s) of size n units, there will be $n!$ ordered samples (si). Let x_{si} $\{s = 1, 2, \dots \binom{N}{n}; i = 1, 2, \dots, M\ (= n!)\}$ be an estimator of a population parameter θ based on the ordered sample (si). Consider a scheme of selection in which the probability of selecting the ordered sample (si) is p_{si}. Then the probability p_s of getting the unordered sample (s) is the sum of the probabilities of getting the ordered samples corresponding to (s)

That is,
$$p_s = \sum_{i=1}^{M} p_{si}.$$

Theorem 1: *If* $\hat{\theta}_0 = x_{si}$ *and* $\hat{\theta}_U = \sum_{i=1} x_{si}\, p'_{si}$ (*where* $p'_{si} = p_{si}/p_s$) *are estimators of the population parameter* θ *then,*

$$\text{(i)} \quad E(\hat{\theta}_U) = E(\hat{\theta}_0)$$

$$\textit{and} \quad \text{(ii)} \quad V(\hat{\theta}_U) \leqslant V(\hat{\theta}_0),$$

where E and V stand for expectation and variance respectively.

Proof :

$$E(\hat{\theta}_U) = \sum_{s=1}^{\binom{N}{n}} \hat{\theta}_U p_s = \sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} x_{si}\, p_{si} = E(\hat{\theta}_0).$$

The variances of the estimators $\hat{\theta}_0$ and $\hat{\theta}_U$ are given by

$$V(\hat{\theta}_0) = \left\{ \sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} x_{si}^2\, p_{si} - \left(\sum_{s=i}^{\binom{N}{n}} \sum_{i=1}^{M} x_{si}\, p_{si} \right)^2 \right\}. \quad \ldots \quad (2.1)$$

$$V(\hat{\theta}_U) = \left\{ \sum_{s=1}^{\binom{N}{n}} \left(\sum_{i=1}^{M} x_{si}\, p'_{si} \right)^2 p_s - \left(\sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} x_{si}\, p_{si} \right)^2 \right\}. \quad \ldots \quad (2.2)$$

Therefore,

$$V(\hat{\theta}_0) - V(\hat{\theta}_U) = \left\{ \sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} \left(x_{si} - \sum_{i=1}^{M} x_{si}\, p'_{si} \right)^2 p_{si} \right\}. \quad \ldots \quad (2.3)$$

This shows that the variance of the unordered estimator $\hat{\theta}_U$ is less than or equal to that of the ordered estimator $\hat{\theta}_0$.

Corollary :

(*i*) *The mean square error of* $\hat{\theta}_U$ *is less than or equal to that of* $\hat{\theta}_0$.

(*ii*) *If* $\hat{V}_0(\hat{\theta}_0) = v_{si}$ *is an ordered estimator of* $V(\hat{\theta}_0)$, *then* $\hat{V}_U(\hat{\theta}_0)$, *an unordered estimator of* $V(\hat{\theta}_0)$, *which has a lesser mean square error than* $\hat{V}_0(\hat{\theta}_0)$ *is given by*

$$\hat{V}_U(\hat{\theta}_0) = \left\{ \sum_{i=1}^{M} v_{si}\, p'_{si} \right\}. \quad \ldots \quad (2.4)$$

(*iii*) *An estimator* $\hat{V}_U(\hat{\theta}_U)$ *of the variance of* $\hat{\theta}_U$ *is given by*

$$\hat{V}_U(\hat{\theta}_U) = \left\{ \hat{\theta}_U^2 - \sum_{i=1}^{M} (x_{si}^2 - v_{si}) p'_{si} \right\}. \quad \ldots \quad (2.5)$$

(*iv*) *The 2k-th moment of* $\hat{\theta}_U$ *is less than or equal to that of* $\hat{\theta}_0$.

That is,

$$E\{\hat{\theta}_U - E(\hat{\theta}_U)\}^{2k} \leqslant E\{\hat{\theta}_0 - E(\hat{\theta}_0)\}^{2k}.$$

It may be noted that an estimator based on $m (m = 2, \ldots, M-1)$ ordered samples drawn from the M ordered samples corresponding to the unordered sample (s) with probability proportional to the sum of their probabilities will be more efficient than the ordered estimator $\hat{\theta}_0$.

3. UNORDERING OF DES RAJ'S ESTIMATORS

Let a sample of size n be drawn from a finite population of N units with varying probabilities without replacement. Suppose the probabilities of selection at the first draw are

$$\{p_j\},\ (j = 1, 2, \ldots, N),\ p_j > 0,\ \sum_{j=1}^{N} p_j = 1. \qquad \ldots \quad (3.1)$$

The scheme of selection of a unit at a particular draw depends on the units already drawn in the sample and not on the order in which they were drawn. For instance, the probabilities of selection at the third draw given that the k-th and l-th units have already been chosen in the first two draws will be given by

$$\left\{ \frac{p_j}{1-p_k-p_l} \right\},\ (j \neq k \neq l).$$

Let $(y_1, y_2, \ldots, y_n)$ and $(p_1, p_2, \ldots, p_n)$ be the values of the units arranged in the order of selection in the sample drawn according to the above scheme and their respective initial probabilities. One of the sets of estimators given by Des Raj in this situation is given by

$$\left.\begin{aligned} x_{si1} &= \left\{ \frac{y_1}{p_1} \right\} \\ x_{si2} &= \left\{ y_1 + \frac{y_2}{p_2}(1-p_1) \right\} \\ x_{sin} &= y_1 + y_2 + \cdots + y_{n-1} + \left(\frac{y_n}{p_n} \right) (1-p_1-p_2-\cdots-p_{n-1} \end{aligned}\right\}. \qquad \ldots \quad (3.2)$$

Each of the above estimators is unbiased for the population total and, therefore,

$$\bar{x}_{si} = \frac{1}{n} \left\{ \sum_{j=1}^{n} x_{sij} \right\}, \qquad \ldots \quad (3.3)$$

is also so. By making use of the fact that x_{sij} and $x_{sij'}$ $(j \neq j')$ are uncorrelated, Des Raj was able to get a non-negative estimator of the variance of $\bar{x}_{si}$ which is given by

$$\hat{V}_0(\bar{x}_{si}) = v_{si} = \frac{1}{n(n-1)} \left\{ \sum_{j=1}^{n} (x_{sij} - \bar{x}_{si})^2 \right\}, \qquad \ldots \quad (3.4)$$

By applying the earlier theorem to $\bar{x}_{si}$ and v_{si} we get more efficient estimators of the population total and the variance of $\bar{x}_{si}$ respectively.

Theorem 2: *Unordering of the ordered estimator,*

$$\hat{\theta}_0 = \sum_{j=1}^{n} c_j x_{sij}, \ \sum_{j=1}^{n} c_j = 1, \ x_{sij} = y_1+y_2+\cdots+y_{j-1}+\frac{y_j}{p_j} 1-p_1-\cdots p_{j-1}), \quad \ldots \quad (3.5)$$

yields an unordered estimator which is independent of the set c_j, $\left\{ \sum_{j=1}^{n} c_j = 1 \right\}$, *namely,*

$$\frac{\sum_{l=1}^{n} y_l \, P(s/l)}{P(s)} , \quad \ldots \quad (3.6)$$

where $P(s/l)$ *is the conditional probability of getting the unordered sample* (s) *given that l-th unit has been selected at the first draw and* $p(s)$ *is the unconditional probability of getting* (s).

Proof: Let $P(s/l_1, l_2, \ldots l_j)$ denote the conditional probability of getting the unordered sample (s) given that l_1-th, l_2-th, …, l_j th units have been selected in the first j draws.

The coefficient of y_l in the estimator got by unordering $\hat{\theta}_0$ in the usual way (Theorem 1) is given by $\frac{1}{P(s)}$ times

$$\sum_{j=1}^{n} c_j \left[\sum_{k=1}^{j-1} \left\{ \sum_{\substack{l_1, l_2, \ldots l_{k-1} \\ l_1 \neq l_2 \neq \ldots l_{k-1} \neq l}} \frac{p_l p_{l_1} p_{l_2} \cdots p_{l_{k-1}}}{(1-p_{l_1})(1-p_{l_1}-p_{l_2}) \ldots (1-p_{l_1}-p_{l_2}-\ldots p_{l_{k-1}}} \times \right. \right.$$

$$\left. \times \ P(s/l \, l_1 \, l_2 \ldots l_{k-1}) \right\}$$

$$+ \sum_{\substack{l_1, l_2, \ldots, l_{j-1} \\ l_1 \neq l_2 \neq \ldots \neq l_{j-1} \neq l}} \frac{p_l \, p_{l_1} \, p_{l_2} \cdots p_{l_{j-1}}}{(1-p_{l_1})(1-p_{l_1}-p_{l_2}) \ldots (1-p_{l_1}-p_{l_2}-\ldots-p_{l_{2-1}})} \times$$

$$\left. \times \ P(s/l \, l_1 \, l_2 \ldots l_{j-1}) \times \frac{1-p_{l_1}-p_{l_2}-\ldots p_{l_{j-1}}}{p_l} \right] \cdot \ldots \quad (3.7)$$

The coefficient of c_1 in the above expression is $P(s/l)$. The theorem will be proved if we show that the coefficient of c_{j+1} is equal to that of c_j in the expression (3.7).

The first $(j-1)$ terms in both the coefficients are the same. The j-th and $(j+1)$th terms in the coefficient of c_{j+1} reduce to the j-th term in the coefficient of c_j because of the equality

$$P(s/l \, l_1 \, l_2 \ldots l_{j-1}) = \sum_{\substack{l_j \\ l_j \neq l_{j-1} \neq \ldots \neq l_1 \neq l}} \frac{p_{l_j}}{(1-p_l-p_{l_1}-\ldots p_{l_{j-1}})} \times P(s/l \, l_1 \, l_2 \ldots, l_j).$$

Therefore,
$$\hat{\theta}_u = \bar{x}_s = \frac{\sum\limits_{l=1}^{n} y_l P(s/l)}{P(s)}.$$

The estimator given in (3.6) is a special case of a general class of estimators considered by Nanjamma, Murthy and Sethi, in an unpublished paper submitted to *Sankhyā*.

Since the unbiased estimators x_{sij} and $x_{sij'}$ $(j \neq j')$ of the population total Y are uncorrelated, we get

$$\sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} (x_{sij}\, x_{sij'})\, p_{si} = Y^2.$$

As this is true for all the $\binom{n}{2}$ pairs of the n estimators.

$$\sum_{s=1}^{\binom{N}{n}} \sum_{i=1}^{M} \Big\{ \sum_{j>j'}^{n} x_{sij}\, x_{sij'} \Big\}\, p_{si} = \binom{n}{2}\, Y^2.$$

Hence an unbiased estimator of Y^2 is given by

$$(\hat{Y^2}) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{M} \Big(\sum_{j>j'}^{n} x_{sij}\, x_{sij'} \Big)\, p'_{si}. \qquad \dots \quad (3.8)$$

From this it follows that an unbiased estimator of the variance of $\bar{x}_s$ is

$$\hat{V}_U(\bar{x}_s) = \bar{x}_s^2 - \frac{1}{\binom{n}{2}} \sum_{i=1}^{M} \Big(\sum_{j>j'}^{n} x_{sij}\, x_{sij'} \Big)\, p'_{si}$$

$$= \frac{1}{[P(s)]^2} \Big[\sum_{l=1}^{n} P(s/l)\{P(s/l)-P(s)\} y_l^2 + 2 \sum_{l>l'}^{n} \{P(s/l)\, P(s/l') - P(s) P(s/ll')\} y_l\, y_{l'}. \Big] \dots \quad (3.9)$$

The following particular cases of equations (3.6) and (3.9) will now be considered.

(i) *Simple random sampling without replacement.* In this case the estimators $\bar{x}_{si}$, $\bar{x}_s$ and their variances are given by

$$\bar{x}_{si} = \frac{1}{n} \Big\{ \sum_{j=1}^{n} (N+n+1-2j) y_j \Big\}, \qquad \dots \quad (3.10)$$

$$\bar{x}_s = \frac{N}{n} \Big\{ \sum_{j=1}^{n} y_j \Big\}, \qquad \dots \quad (3.11)$$

$$V(\bar{x}_{si}) = \frac{\sigma^2}{n} \Big\{ N(N-n) + \frac{n^2-1}{3} \Big\}, \qquad \dots \quad (3.12)$$

and
$$V(\bar{x}_s) = \frac{\sigma^2}{n} N(N-n), \quad \ldots \quad (3.13)$$

where
$$\sigma^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N} (y_i - \bar{y})^2\right\}.$$

Therefore,
$$V(\bar{x}_{si}) - V(\bar{x}_s) = \frac{n^2-1}{3n}\,\sigma^2. \quad \ldots \quad (3.14)$$

Comparison of the above expressions for the variances shows that $V(\bar{x}_s) \leqslant V(\bar{x}_{si})$. This is otherwise obvious also as $N\bar{y}$ is known to be the best unbiased linear estimator of Y. It is interesting to note that the variance estimator of $\bar{x}_s$ given in equation (3.9) reduces to the estimator commonly used, namely,

$$\hat{V}_U(\bar{x}_s) = N(N-n)\,\frac{s^2}{n}, \quad \ldots \quad (3.15)$$

where
$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} (y_i - \bar{y})^2\right\}.$$

It may be pointed out that the divergence of Des Raj's estimator from the best unbiased linear estimator led to a search for a more efficient estimator.[1]

(ii) *Sampling of two units with varying probabilities without replacement.* This case is of importance as in actual practice one will, in general, be choosing two units from each stratum in stratified sampling. In this case $\bar{x}_{si}$ and $\bar{x}_s$ are given by

$$\bar{x}_{si} = \tfrac{1}{2}\left\{(1+p_1)\frac{y_1}{p_1} + (1-p_1)\frac{y_2}{p_2}\right\}, \quad \ldots \quad (3.16)$$

and
$$\bar{x}_s = \frac{1}{2-p_1-p_2}\left\{(1-p_2)\frac{y_1}{p_1} + (1-p_1)\frac{y_2}{p_2}\right\}. \quad \ldots \quad (3.17)$$

The sampling variances of these two estimators are given by

$$V(\bar{x}_{si}) = \tfrac{1}{4}\left\{\sum_{s=1}^{\binom{N}{2}} p_1 p_2 (2-p_1-p_2)\left(\frac{y_1}{p_1} - \frac{y_2}{p_2}\right)^2\right\}, \quad \ldots \quad (3.18)$$

and
$$V(\bar{x}_s) = \left\{\sum_{s=1}^{\binom{N}{2}} p_1 p_2 \frac{(1-p_1-p_2)}{(2-p_1-p_2)}\left(\frac{y_1}{p_1} - \frac{y_2}{p_2}\right)^2\right\}. \quad \ldots \quad (3.19)$$

---

[1] Lahiri conjectured that Des Raj's estimators can be improved by weighting the different ordered estimators by their respective probabilities and in fact suggested the estimator given in (3.17)

Therefore, $$V(\bar{x}_{si})-V(\bar{x}_s) = \tfrac{1}{4}\left\{\sum_{s=1}^{\binom{N}{2}} p_1\,p_2\,\frac{(p_1+p_2)^2}{(2-p_1-p_2)}\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2\right\}. \qquad \ldots \quad (3.20)$$

In the case of simple random sampling without replacement the expression (3.18) above becomes

$$V(\bar{x}_{si})-V(\bar{x}_s) = \frac{\sigma^2}{2},$$

which is a particular case of the expression (3.12).

The estimator of variance of $\bar{x}_{si}$ given by Des Raj is,

$$\hat{V}_0(\bar{x}_{si}) = v_{si} = \tfrac{1}{4}\,(1-p_1)^2\,\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2. \qquad \ldots \quad (3.21)$$

By applying the earlier theorem to this v_{si}, we get a more efficient estimator of the variance, namely,

$$\hat{V}_U\,(\bar{x}_{si}) = \tfrac{1}{4}\,(1-p_1)(1-p_2)\,\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2, \qquad \ldots \quad (3.22)$$

and this also is non-negative. Substituting the relevant values in equation (3·7), we get an unbiased estimator of the variance of $\bar{x}_s$, namely,

$$\hat{V}_U(\bar{x}_s) = \frac{(1-p_1)(1-p_2)(1-p_1-p_2)}{(2-p_1-p_2)^2}\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2, \qquad \ldots \quad (3.23)$$

which is always non-negative.

4. Unordering of Das' estimators

With the notation adopted in section 3, the set of estimators proposed by Das is given by,

$$x'_{si1} = \left(\frac{y_1}{p_1}\right)$$

$$x'_{si2} = \frac{1}{(N-1)}\,\frac{(1-p_1)}{p_1}\left(\frac{y_2}{p_2}\right)$$

$$x'_{sin} = \frac{(1-p_1)(1-p_1-p_2)\ldots(1-p_1-p_2-\ldots-p_{n-1})}{(N-1)(N-2)\ldots(N-n+1)p_1\,p_2\cdots p_{n-1}}\left(\frac{y_n}{p_n}\right). \qquad \ldots \quad (4.1)$$

Each of the above estimators is unbiased for the population total Y and so is their mean

$$\bar{x}'_{si} = \frac{1}{n}\sum_{j=1}^{n} x'_{sij}. \qquad \ldots \quad (4.2)$$

An unbiased estimator of the variance of $\bar{x}'_{si}$ is

$$\hat{V}_0(\bar{x}'_{si}) = v'_{si} = \bar{x}'^{2}_{si} - \left\{ \frac{1}{n} \sum_{j=1}^{n} \left(x'_{sij}\, y_j \right) + \frac{N-1}{\binom{n}{2}} \sum_{j>'=1}^{n} (x'_{sij}\, y_{j'}) \right\}. \quad \dots \quad (4.3)$$

In sampling the first unit with varying probability and the remaining $(n-1)$ units from $(N-1)$ units with equal probability without replacement, the estimator $\bar{x}'_{si}$ becomes

$$\bar{x}'_{si} = \frac{1}{n} \left\{ \frac{\sum\limits_{j=1}^{n} y_j}{p_1} \right\}. \quad \dots \quad (4.4)$$

Applying the Theorem 1 to this $\bar{x}'_{si}$, we get the unordered estimator

$$\bar{x}'_{s} = \left\{ \frac{\sum\limits_{j=1}^{n} y_j}{\sum\limits_{j=1}^{n} p_j} \right\}. \quad \dots \quad (4.5)$$

This shows that the above unbiased ratio estimator is more efficient than the estimator given by Das. Lahiri (1951) and Midjuno (1952) have given sampling procedures which lead to the above estimator.

Substituting the relevant values in equation (2.5) we get an unbiased variance estimator of $\bar{x}'_s$, namely,

$$\hat{V}_U(\bar{x}'_s) = \bar{x}'^{2}_{s} - \left\{ (N-1) \frac{\left(\sum\limits_{j=1}^{n} y_j\right)^2 - (N-n)\left(\sum\limits_{j=1}^{n} y_j^2\right)}{(n-1)\left(\sum\limits_{j=1}^{n} p_j\right)} \right\}. \quad \dots \quad (4.6)$$

5. NUMERICAL EXAMPLES

To study the relative performance of the ordered and unordered estimators, the following population given by Yates and Grundy (1953) will be considered.

unit	p	y	y/p
1	.1	0.5	5
2	.2	1.2	6
3	.3	2.1	7
4	.4	3.2	8
total	1.0	7.0	

This was deliberately chosen by them as being more extreme than will normally be encountered in practice. The object is to estimate the population total by selecting two units. Two schemes of selection will be considered:

Case (i): first unit with varying probability and second unit with equal probability without replacement;

Case (ii): both the units with varying probabilities without replacement.

For the prupose of comparison, the estimator and the variance estimator proposed by Horvitz and Thompson (1952) together with the variance estimator given by Yates and Grundy are also considered.

For the sake of convenience, the expressions for the different estimators given in subsequent tables are given below. Suppose (y_1, y_2) is the ordered sample drawn from a finite population of N units.

Case (i):

$$\bar{x}_{si} = \frac{1}{2}\left\{ (1+p_1)\left(\frac{y_1}{p_1}\right) + (N-1)y_2 \right\}. \quad \ldots \ (5.1)$$

$$\hat{V}_0(\bar{x}_{si}) = v_{si} = \frac{1}{4}\left\{(1-p_1)\left(\frac{y_1}{p_1}\right) - (N-1)y_2\right\}^2. \quad \ldots \ (5.2)$$

$$\hat{V}_U(\bar{x}_{si}) = \frac{v_{s1}\, p_1 + v_{s2}\, p_2}{p_1+p_2}. \quad \ldots \ (5.3)$$

$$\bar{x}_s = \frac{1}{2(p_1+p_2)}\left[\left\{ 1+p_1+(N-1)p_2 \right\} y_1 + \left\{1+p_2+(N-1)p_1\right\} y_2\right]. \quad (5.4)$$

$$\hat{V}_U(\bar{x}_s) = \hat{V}_U(\bar{x}_{si}) - \frac{(\bar{x}_{s1}-\bar{x}_s)^2\, p_1 + (\bar{x}_{s2}-\bar{x}_s)^2 p_2}{p_1+p_2}. \quad \ldots \ (5.5)$$

$$\bar{x}'_{si} = \frac{y_1+y_2}{2p_1}. \quad \ldots \ (5.6)$$

$$\hat{V}_0(\bar{x}'_{si}) = v'_{si} = \bar{x}'^2_{si} - \frac{1}{2p_1}\left\{y_1^2+y_2^2+2(N-1)y_1y_2 \right\}. \quad \ldots \ (5.7)$$

$$\hat{V}_U(\bar{x}'_{si}) = \frac{v'_{si}\, p_1 + v'_{s2}\, p_2}{p_1+p_2}. \quad \ldots \ (5.8)$$

$$\bar{x}_s = \left\{ \frac{y_1+y_2}{p_1+p_2} \right\}. \quad \ldots \ (5.9)$$

$$\hat{V}_U(\bar{x}'_s) = \bar{x}'^2_s - \left\{\frac{y_1^2+y_2^2+2\,(N-1)\,y_1\,y_2}{p_1+p_2}\right\}. \quad \ldots \ (5.10)$$

$$\bar{y}_{HT} = \left(\frac{y_1}{\pi_1}\right) + \left(\frac{y_2}{\pi_2}\right) \text{ where } \pi_1 = \frac{1}{N-1}\,\{(N-2)p_1+1\}$$

and

$$\pi_2 = \frac{1}{N-1}\,\{(N-2)p_2+1\}. \quad \ldots \ (5.11)$$

$$\hat{V}_{HT}\,(\bar{y}_{HT}) = (1-\pi_1)\left(\frac{y_1^2}{\pi_1^2}\right)+(1-\pi_2)\left(\frac{y_2^2}{\pi_2^2}\right)+2\;\frac{\pi_{12}-\pi_1\pi_2}{\pi_{12}}\;\frac{y_1\,y_2}{\pi_1\,\pi_2},$$

where

$$\pi_{12} = \frac{p_1+p_2}{(N-1)}. \qquad \dots \quad (5.12)$$

$$\hat{V}_{YG}\,(\bar{y}_{HT}) = \frac{\pi_1\pi_2-\pi_{12}}{\pi_{12}}\left(\frac{y_1}{\pi_1}-\frac{y_2}{\pi_2}\right)^2. \qquad \dots \quad (5.13)$$

Case (ii):

$$\bar{x}_{si} = \tfrac{1}{2}\left\{(1+p_1)\left(\frac{y_1}{p_1}\right)+(1-p_1)\left(\frac{y_2}{p_2}\right)\right\}.$$

$$\hat{V}_0(\bar{x}_{si}) = \tfrac{1}{4}\,(1-p_1)^2\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2.$$

$$\hat{V}_U\,(\bar{x}_{si}) = \tfrac{1}{4}\,(1-p_1)\,(1-p_2)\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2.$$

$$\bar{x}_s = \frac{1}{2-p_1-p_2}\left\{(1-p_2)\left(\frac{y_1}{p_1}\right)+(1-p_1)\left(\frac{y_2}{p_2}\right)\right\}.$$

$$\hat{V}_U\,(\bar{x}_s) = \frac{(1-p_1)(1-p_2)(1-p_1-p_2)}{(2-p_1-p_2)^2}\left(\frac{y_1}{p_1}-\frac{y_2}{p_2}\right)^2.$$

$$\bar{x}'_{si} = \tfrac{1}{2}\left\{\left(\frac{y_1}{p_1}\right)+\frac{1}{N-1}\left(\frac{1-p_1}{p_1}\right)\left(\frac{y_2}{p_2}\right)\right\}. \qquad \dots \quad (5.14)$$

$$\hat{V}_0\,(\bar{x}'_{si}) = v'_{si} = \bar{x}'^2_{si} - \tfrac{1}{2}\left\{\left(\frac{y_1^2}{p_1}\right)+\frac{1}{N-1}\,\frac{(1-p_1)}{p_1}\left(\frac{y_2^2}{p_2}\right)+\frac{2(1-p_1)}{p_1}\;\frac{y_1\,y_2}{p_2}\right\}. \quad (5.15)$$

$$\hat{V}_U(\bar{x}'_s) = \frac{1}{2-p_1-p_2}\left\{(1-p_2)v'_{s1}+(1-p_1)v'_{s2}\right\}. \qquad \dots \quad (5.16)$$

$$\bar{x}'_s = \frac{1}{2-p_1-p_2}\left\{(1-p_2)\bar{x}'_{s1}+(1-p_1)\bar{x}'_{s2}\right\}. \qquad \dots \quad (5.17)$$

UNBIASED ESTIMATE OF ERROR VARIANCE CASE (i)

sample	$\hat{V}_0(\bar{x}_{si})$	$\hat{V}_U(\bar{x}_{si})$	$\hat{V}_U(\bar{x}_s)$	$\hat{V}_0(\bar{x}'_{si})$	$\hat{V}_U(\bar{x}'_{si})$	$\hat{V}_U(\bar{x}'_s)$	$\hat{V}_{HT}(\bar{y}_{HT})$	$\hat{V}_{YG}(\bar{y}_{HT})$
12	.20	1.88	1.87	45.80	18.49	14.48	−1.11	1.51
13	.81	2.37	2.30	114.20	28.93	14.85	2.27	4.33
14	6.50	3.48	3.25	241.80	45.38	14.58	6.45	7.34
21	2.72	1.88	1.87	4.84	18.49	14.48	−1.11	1.51
23	.56	.48	.44	15.64	3.44	1.62	.77	.92
24	5.76	2.16	1.94	34.20	2.63	−4.09	3.93	3.05
31	2.89	2.37	2.30	.51	28.93	14.85	2.27	4.33
32	.42	.48	.44	−4.70	3.44	1.62	.77	.92
34	5.52	2.69	2.60	−13.59	−20.01	−21.20	3.01	.72
41	2.72	3.48	3.25	−3.72	45.38	14.58	6.45	7.34
42	.36	2.16	1.94	−13.15	2.63	−4.09	3.93	3.05
43	.56	2.69	2.60	−24.82	−20.01	−21.20	3.01	.72
true error variance	2.223	2.223	2.103	9.701	9.701	0.363	2.884	2.884
variance of estimated error variance	1.9912	.8357	.7543	2583.03	491.72	194.33	4.7317	5.5115

UNBIASED ESTIMATE OF ERROR VARIANCE CASE (ii)

sample	$\hat{V}_0(\bar{x}_{si})$	$\hat{V}_U(\bar{x}_{si})$	$\hat{V}_U(\bar{x}_s)$	$\hat{V}_0(\bar{x}'_{si})$	$\hat{V}_U(\bar{x}'_{si})$	$\hat{V}_U(\bar{x}'_s)$	$\hat{V}_{HT}(\bar{\ }_{HT})$	$\hat{V}_{YG}(\bar{y}_{HT})$
12	.20	.18	.17	93.20	49.60	42.95	−6.20	.41
13	.81	.63	.59	114.20	48.17	34.12	−4.70	1.52
14	1.82	1.22	1.08	134.60	47.92	27.39	−.58	2.79
21	.16	.18	.17	10.84	49.60	42.95	−6.20	.41
23	.16	.14	.12	11.78	2.55	1.71	−3.79	.36
24	.64	.48	.39	10.38	−3.07	−5.03	1.21	1.08
31	.49	.63	.59	−3.18	48.17	34.12	−4.70	1.52
32	.12	.14	.12	−5.52	2.55	1.71	−3.79	.36
34	.12	.10	.07	−12.80	−15.07	−15.25	5.02	.18
41	.81	1.22	1.08	−9.86	47.92	27.39	−.58	2.79
42	.36	.48	.39	−13.15	−3.07	−5.03	1.21	1.08
43	.09	.10	.07	−17.01	−15.07	−15.25	5.02	.18
true error variance	0.365	0.365	0.312	5.435	5.435	1.107	0.823	0.823
variance of estimated error variance	.1575	.1236	.1004	1542.49	606.18	350.73	14.8692	.6811

$$\hat{V}_U(\bar{x}'_s) = \hat{V}_U(\bar{x}'_{si}) - \frac{1}{2-p_1-p_2}\{(1-p_2)(\bar{x}'_{s1}-\bar{x}'_s)^2+(1-p_1)(\bar{x}'_{s2}-\bar{x}'_s)^2\}. \quad \ldots \quad (5.18)$$

$$\bar{y}_{HT} = \left(\frac{y_1}{\pi_1}\right) + \left(\frac{y_2}{\pi_2}\right) \text{ where } \pi_1 = p_1\left\{1+\sum_{j\neq 1}^{N}\left(\frac{p_j}{1-p_j}\right)\right\},$$

and

$$\pi_2 = p_2\left\{1+\sum_{j\neq 2}^{N}\left(\frac{p_j}{1-p_j}\right)\right\}. \quad \ldots \quad (5.19)$$

$$\hat{V}_{HT}(\bar{y}_{HT}) = (1-\pi_1)\left(\frac{y_1^2}{\pi_1^2}\right) + (1-\pi_2)\left(\frac{y_2^2}{\pi_2^2}\right)^2 + \frac{\pi_{12}-\pi_1\pi_2}{\pi_{12}}\left(\frac{y_1}{\pi_1}\right)\left(\frac{y_2}{\pi_2}\right),$$

where

$$\pi_{12} = \left\{\frac{p_1\,p_2(2-p_1-p_2)}{(1-p_1)(1-p_2)}\right\},$$

$$\hat{V}_{YG}(\bar{y}_{HT}) = \frac{\pi_1\,\pi_2-\pi_{12}}{\pi_{12}}\left(\frac{y_1}{\pi_1}-\frac{y_2}{\pi_2}\right)^2. \quad \ldots \quad (5.20)$$

The results given in the above table show that for this population unordering of Das' estimators in the above two cases yields estimators which are much more efficient than the corresponding ordered estimators. Of the three unordered unbiased estimators of the population total, namely, $\bar{x}_s$, $\bar{x}'_s$ and $\bar{y}_{HT}$, $\bar{x}'_s$ in case (i) and $\bar{x}_s$ in case (ii) have the least variance. It can also be seen that it is possible to improve substantially on the ordered variance estimators by unordering them.

The author is grateful to Prof. D. B. Lahiri for his valuable suggestions in preparing the paper.

REFERENCES

DAS, A. C. (1951): Two phase sampling and sampling with varying probabilities. *Bull. Int. Stat. Inst.*, **33**, 105–112.

DES RAJ (1956): Some estimators in sampling with varying probabilities without replacement. *J. Amer. Stat. Ass.*, **51**, 274, 269-284.

HORVITZ, D. G. & THOMPSON, D. I. (1952): A generalisation of sampling without replacement from a finite universe. *J. Amer. Stat. Ass.*, **47**, 663-85.

LAHIRI, D. B. (1951): A method of sample selection providing unbiased ratio estimates. *Bull. Int. Stat. Inst.*, **33**, 133–140.

MIDZUNO, H. (1952): On the sampling system with probability proportional to the sum of sizes. *Ann. Inst. Stat. Math.*, (Japan), **3**, 99–108.

ROY CHOWDHURY, D. K. (1956): Sampling with varying probabilities (part of thesis submitted for the Associationship of the Indian Statistical Institute).

YATES, F. AND GRUNDY, P. M. (1953): Selection without replacement from within strata with probability proportional to size. *J. Roy. Stat. Soc.*, B, **15**, 253-61.

Paper received : April, 1957.

A LOWER BOUND TO THE PROBABILITY OF STUDENT'S RATIO

By SAIBAL KUMAR BANERJEE
Indian Statistical Institute, Calcutta

If the parent population is normal, Student's ratio $\frac{\bar{x}-m}{s/\sqrt{n}}$ follows the t distribution

$$\frac{1}{\sqrt{n-1}} \cdot \frac{\left|\overline{\frac{n}{2}}\right.}{\left|\overline{\frac{1}{2}}\right.\left|\overline{\frac{n-1}{2}}\right.} \cdot \left(1+\frac{t^2}{n-1} \right)^{-\frac{n}{2}} dt.$$

An examination of the behaviour of Student's ratio for samples drawn from non-normal populations was made by Bartlett (1935), Geary (1936), and Gayen (1949). From consideration of estimation by confidence interval, however, a lower bound to the probability of the event $\bar{x}+\frac{ts}{\sqrt{n}} \geqslant m \geqslant \bar{x}-\frac{ts}{\sqrt{n}}$ may be of some practical value. Starting from the approach of Tshebysheff's lemma it is possible to derive an expression for the lower bound to the probability of such an event. The expression for the lower bound as worked out, involves t, n and B_2-coefficient of the parent population and applies to samples drawn from any parent population, continuous or discontinuous, having finite fourth cumulant. The proof rests upon a simple lemma.

Lemma : *Let x be a variable with mean m and variance σ^2. The probability $P(k^2)$ of the inequality $x \geqslant m-k^2$, satisfy the relation*

$$P(k^2) \geqslant \frac{k^4}{\sigma^2+k^4}, \qquad \ldots \quad (1)$$

whatever k^2 may be.

Let $x_1, x_2 \ldots, x_n$ be a sample of size n drawn at random with replacement from a population with mean, second cumulant and fourth cumulant denoted respectively by m, k_2 and k_4. Let a variate u be defined as

$$u = \frac{t^2 s^2}{n} -(\bar{x}-m)^2, \qquad \ldots \quad (2)$$

where

$$\bar{x} = \frac{\sum_{1}^{n} x_i}{n},$$

$$s^2 = \frac{1}{n-1} \left\{ \sum_{1}^{n} x_i^2 - n\bar{x}^2 \right\} \text{ and } t^2 > 1.$$

From (1) we have

$$\text{prob}\left\{ u \geqslant 0 \right\} \geqslant \frac{\{E(u)\}^2}{V(u)+\{E(u)\}^2} \geqslant \frac{\{E(u)\}^2}{E(u^2)}. \qquad \ldots \quad (3)$$

We have

$$E(u) = (t^2-1)\,\frac{k_2}{n} \qquad \ldots \quad (4)$$

$$E(u^2) = E\left[\frac{t^4s^4}{n^2} + (\bar{x}-m)^4 - \frac{2t^2s^2}{n}(\bar{x}-m)^2 \right]. \qquad \ldots \quad (5)$$

Since

$$\left.\begin{aligned} E(s^4) &= \frac{k_4}{n} + \frac{2k_2^2}{n-1} + k_2^2 \\ E(\bar{x}-m)^4 &= \frac{k_4}{n^3} + \frac{3k_2^2}{n^2} \\ E\{s^2(\bar{x}-m)^2\} &= \frac{k_4}{n^2} + \frac{k_2^2}{n}. \end{aligned}\right\} \qquad \ldots \quad (6)$$

$$E(u^2) = \frac{t^4}{n^2}\left(\frac{k_4}{n} + \frac{2k_2^2}{n-1} + k_2^2 \right) + \frac{k_4}{n^3} + \frac{3k_2^2}{n^2} - \frac{2t^2}{n}\left(\frac{k_4}{n^2} + \frac{k_2^2}{n} \right)$$

$$= \frac{k_4}{n^3}(t^2-1)^2 + \frac{k_2^2}{n^2}(t^2-1)^2 + \frac{2k_2^2}{n^2}\left\{ \frac{t^4}{n-1} + 1 \right\}$$

$$= \frac{k_2^2(t^2-1)^2}{n^2}\left[\frac{B_2-3}{n} + 1 + \frac{2}{(t^2-1)^2}\left\{ \frac{t^4}{n-1} + 1 \right\} \right], \qquad \ldots \quad (7)$$

where $B_2 = \frac{k_4}{k_2^2} + 3.$

From (3), (4) and (7) we have

$$\text{prob}\,\{u \geqslant 0\} \geqslant \frac{1}{\frac{B_2-3}{n} + 1 + \frac{2}{(t^2-1)^2}\left\{ \frac{t^4}{n-1} + 1 \right\}}. \qquad \ldots \quad (8)$$

Since $u = \frac{t^2s^2}{n} - (\bar{x}-m)^2$, from (8) we have,

$$\text{prob}\left\{ \bar{x} + \frac{ts}{\sqrt{n}} \geqslant m \geqslant \bar{x} - \frac{ts}{\sqrt{n}} \right\} \geqslant \frac{1}{\frac{B_2-3}{n} + 1 + \frac{2}{(t^2-1)^2}\left\{ \frac{t^4}{n-1} + 1 \right\}}. \qquad \ldots (9)$$

The expression for the lower bound of the probability of the inequality $\bar{x}+\frac{ts}{\sqrt{n}} \geqslant m \geqslant \bar{x}-\frac{ts}{\sqrt{n}}$ as derived in (9) depends upon t, n and B_2 coefficient of the parent population. Numerical values of the lower bound of the probability for $t = 3$, $n = 4, 6, 8, 10, 12, 20$ and 30, and $B_2=1, 2, 3, 4, 5, 6, 8$, and 10 are given in the following Table.

LOWER BOUND OF PROBABILITY OF THE INEQUALITY $\bar{x}+\frac{3s}{\sqrt{n}} \geqslant m \geqslant \bar{x}-\frac{3s}{\sqrt{n}}$

(values worked out from expression (9) taking $t = 3$)

B_2 coefficient of the parent population	sample size						
	4	6	8	10	12	20	30
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.727	0.830	0.875	0.899	0.914	0.939	0.951
2	0.615	0.729	0.789	0.825	0.849	0.897	0.921
3	0.533	0.650	0.718	0.762	0.793	0.859	0.894
4	0.471	0.587	0.659	0.708	0.744	0.823	0.868
5	0.421	0.535	0.609	0.661	0.700	0.791	0.844
6	0.381	0.491	0.566	0.620	0.662	0.761	0.821
8	0.320	0.422	0.496	0.552	0.596	0.707	0.778
10	0.276	0.370	0.441	0.497	0.542	0.660	0.740

From the Table it is seen that if $n = 8$, probability of the inequality $\bar{x}+\frac{3s}{\sqrt{n}} \geqslant m \geqslant \bar{x}-\frac{3s}{\sqrt{n}}$ is greater than .60, even if B_2 be as high as 5.0. If, however, B_2 be 4.0 or less, the probability of the inequality exceeds or equals 0.659. It is further seen from the Table that if $n \geqslant 12$ confidence statement regarding the population mean in the form $\bar{x} \pm \frac{3s}{\sqrt{n}}$ may be made with a confidence coefficient of the order of 0.700 if B_2 be less than or equal to 5.0.

The expression for the lower bound suggests some lines of investigation and these are being investigated. In this connection it may be stated with certain amount of reservation, that if prior knowledge of B_2 or rather an upper bound of the numerical value of B_2 coefficient of the parent population is available, it may be possible to test a suggested value m_0 for the sample mean, on the basis of a sample of size n and sample values $x_1, x_2 \ldots, x_n$.

REFERENCES

BARTLETT, M. S. (1935): The effect of non-normality on the t-distribution. *Proc. Camb. Phil. Soc.*, **31**, 223–231.

GEARY, R. C. (1936): The distribution of student's ratio for non-normal sample. *Supp. J. Roy. Stat. Soc.*, 3, 178–184.

GAYEN, A. K. (1949): Distribution of student's t is random samples of any size drawn from non-normal universes. *Biometrika*, **36**, 353–369.

USPENSKY, J. V. (1937): *Introduction to Mathematical Probability*. McGraw Hill Book Company.

Paper received : Ausust, 1956.

BUREAU OF EDNL. & [illegible] RESEARCH : : LIBRARY

SANKHYĀ : THE INDIAN JOURNAL OF STATISTICS

1. Place of Publication : 204/1 Barrackpore Trunk Road, Calcutta-35

2. Periodicity of its Publication : Quarterly

3. Printer's Name : Kalipada Mukherjee
Nationality : Indian
Address : 204/1 Barrackpore Trunk Road, Calcutta-35

4. Publisher's Name : Prafulla Chandra Mahalanobis
Nationality : Indian
Address : 204/1 Barrackpore Trunk Road, Calcutta-35

5. Editor's Name : P. C. Mahalanobis
Nationality : Indian
Address : 204 Barrackpore Trunk Road, Calcutta-35

6. Names and addresses of individuals who own the newspaper and partners or shareholders holding more than one per cent of the total capital : Published by Statistical Publishing Society
(non-profit distributing agency)

I, Prafulla Chandra Mahalanobis, hereby declare that the particulars given above are true to the best of my knowledge and belief.

Date.......March....1957

Sd/ Prafulla Chandra Mahalanobis